DOCUMENT RESUME

ED 181 058 TM 009 966

AUTHOR	Ingle, Robert B., Ed.: And Others
TITLE	Assessment of Student Competence in the Public Schools.
INSTITUTION	Phi Delta Kappa, Bloomington, Ind. Center on Evaluation and Research.
PUB DATE	[78]
NOTE	141p.: Papers presented at three meetings of the National Symposium for Professionals in Evaluation and Research (NSPER) (Cincinnati, OH: Kansas City, MO: and Portland, OR: 1978)
AVAILABLE FROM	Phi Delta Kappa, P.O. Box 789, Bloomington, IN 47402 ($3.50)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Basic Skills: Competence: Conference Reports: Educational Assessment: Elementary Secondary Education: *Minimum Competency Testing: Performance Based Education: Program Descriptions: Program Development: *School Districts: *State Programs: Student Testing: *Testing Problems: *Testing Programs: Test Results
IDENTIFIERS	Colorado (Denver): Denver Public Schools CO: Florida: Kanawha County Schools WV: Missouri: Oregon: Oregon (Portland): Portland School District OR

ABSTRACT

Three conference sessions focused on attempts to implement minimum competency testing programs at various educational levels. In each of the sessions particular attention was paid to planning for the development of a minimum competency testing program: the processes of development, initiation, and interpretation of the results: and difficulties encountered at each stage. In addition, consideration was given to questions about the best measuring techniques and the advantages and disadvantages of various approaches to minimum competency test construction that have been used or suggested. Each session included case studies of a school system and a state effort to implement a minimum competency measurement program. A presentation by a measurement specialist was also made at each session. The complexity of minimum competency testing was brought out in the various case studies which provided insight into the realities of implementing a program in an area in which so many segments of the public have very strong opinions. Testing programs were shown to force professional educators, schools, and the public to make some difficult decisions about what competency is: what a diploma means: what should be done to best educate students: and how minimum competency testing programs may or may not meet community and student needs. (Author/GDC)

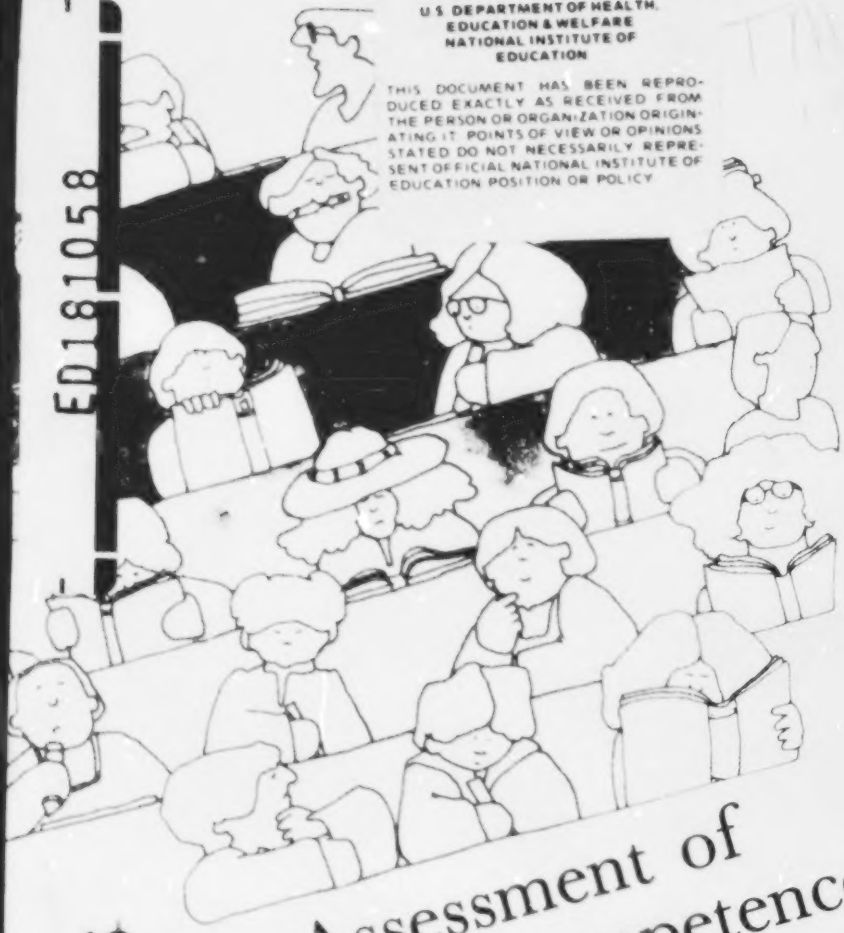
****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED181058

TM009 966

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY



Assessment of Student Competence

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Kliewer

PDK

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A CEDR Monograph
PHI DELTA KAPPA

2

THE ASSESSMENT OF STUDENT COMPETENCE IN THE PUBLIC SCHOOLS

Edited by

ROBERT B. INGLE
University of Wisconsin-Milwaukee

MARY R. CARROLL
Phi Delta Kappa

WILLIAM J. GEPHART
Phi Delta Kappa

Preface

One of the issues about which Phi Delta Kappans have expressed special concern and interest in the last two years is minimum competency testing. The Center on Evaluation, Development, and Research of Phi Delta Kappa focused the annual sessions of the National Symposium for Professionals in Evaluation and Research (NSPER) on attempts to implement minimum competency testing programs at different educational levels.

In each of the three sessions particular attention was paid to planning for the development of a minimum competency testing program, the processes of development, initiation, and interpretation of the results, and difficulties encountered at each stage. In addition consideration was given to questions about the best measuring techniques and the advantages and disadvantages of various approaches to minimum competency test construction that have been used or suggested. Each session included case studies of a school system and a state effort to implement a minimum competency measurement program. A presentation by a measurement specialist was also made at each session.

Beginning with the sessions in Cincinnati, Ohio and continuing through those in Kansas City, Missouri and Portland, Oregon, it was apparent that the issue of minimum competency testing is extremely complex. This was brought out in the various case studies which provided insight into the realities of implementing a program in an area in which so many segments of the public have very strong opinions. The trade-offs in test focus and construction that must be made, often arbitrarily, are brought out both in some case studies and in the papers on measurement.

Finally, the report on the discussion and recommendations of each group of participants serves to reinforce the view presented in most of the papers that minimum competency testing is a program that forces professional educators, the schools, and the public to make some difficult decisions about what competency is; what a diploma means; what should be done to best educate students; and how minimum competency testing programs may or may not meet the needs of students and the community at large.

Plan for the NSPER: 78 Sessions

In the May 1978 issue of the *Kappan*, devoted to MINIMUM COMPETENCY ASSESSMENT, Chris Pipho, Associate Director for Research and Information, Education Commission of the States, said,

> . . . it (minimum competency testing) could only be described as "fast moving." Sweeping concepts were included in one-page bills; state board mandates were often short paragraphs asking local districts or the staff of the state department of education to "implement a full-scale program," sometimes very quickly. Knowledge of criterion-referenced and norm-references tests was often in short supply, and mandates were drafted without much attention to information on existing testing or competency-based programs.

As the rush to get on with it developed (33 states had some effort underway by mid-March 1978), questions and problems surfaced. In that same *Kappan* Henry M. Brickell listed seven: What competencies? How to measure? When to measure? One minimum or many? How high the minimum? Minimums for students or schools? And, what to do with the incompetent? To this list we add the question of who. Who selects the competencies to be measured? Who sets the minimums? Who determines how to measure them? Clearly competency assessment is a problematic area for educators.

BUT, we are in it! And thus, NSPER: 78 will focus on competency assessment. The objective for NSPER: 78 is to develop answers to the above questions using knowledge of the measurement process, experience from selected state and local competency assessment efforts, and the collected expertise of the NSPER: 78 participants.

Three NSPER: 78 sessions are planned. Each will have the same theme and schedule. The cast of characters will be different. A report, covering the presentations and discussions at the three NSPER: 78 sessions will be produced.

Each NSPER: 78 session will have case study presentations of programs at the state and local level and a presentation about measurement technology related to competency assessment. These presentations, small group discussions, and a plenary roundtable will focus on answers to the questions above.

As always NSPER: 78 provides an opportunity for participant exchange of information among people with similar interests, responsibilities, and problems.

The ten questions posed in the announcement were expanded for use in the sessions. They served as an organizing framework for the papers presented at each conference. They were also used by the participants in the small discussion groups. The expanded questions were the following.

1. What competencies? What is a competence and is there a set of them we propose to assess? Assessment (measurement) is a process that assumes concreteness, observability, and stability across instances. Do "competencies" have the concreteness that is demanded by and for measurement? Or, are they "warm fuzzies" that give us lots of wiggle room and measurement fits?

2. Who selects the competencies to be measured? The learner, his family, the teacher, school, state, and nation all seem to act as if they have a stake in the game at times AND as if it is controlled elsewhere at others. Who should select, everyone and no one?

3. How to measure the competencies? Objective, standardized, actuarially based procedures applied to performance, via subjective procedures attached to performance, or both? Obtrusiveness vs. unobtrusiveness? Norm referenced, criterion referenced, domain referenced?

4. Who determines how to measure? The "measurement industry," the educational practitioner, the policy makers, the lay public, and others have ideas about the how to measure question and the credibility of the different ways to measure minimum competency. Who decides? By what process will these decisions be reached?

5. Who sets the minimums? Who has the responsibility, the authority, the expertise, the knowledge of the world today and tomorrow to set the minimum on a competence?

6. How high the minimum? 100 percent, 80, 50? Absolute levels or relative levels? Is this an accountability concern or a concern for facilitating growth?

7. Minimums for students or schools? Are schools being asked to standardize capability, to build a floor of things each person is capable of doing? Is competency-based education a concern for the individual and his or her idiosyncrasies?

8. One minimum or many? Are we setting a bar over which everyone must jump? Or are these alternatives in a competence?

9. When to measure? Early, late, frequently? At one point or on demand?

10. What to do with the incompetent? Program for them as an expected, natural, and accepted artifact of the measurement? Act as if they do not exist? Categorize and stereotype as undesirable and program correctives for them?

CONTENTS

The Plan for the NSPER: 78 Sessions..... ii

The Cincinnati, Ohio Session

Competency Measurement at the Local Level: A Case Study of Kanawha County Schools, West Virginia.... 1
Catherine Candor-Chandler

The Assessment of Student Competencies in Florida's Public Schools..... 24
Thomas H. Fisher

The Technology of Competency Measurement..... 39
William A. Mehrens (Michigan State University)

The Kansas City, Missouri Session

Competency Measurement at the Local Level (The Denver, Colorado Public Schools)..... 59
Barry B. Beal

A Case Study of Basic Skills Testing by the Missouri Department of Elementary and Secondary Education..... 68
Charles Foster

Basic Issues Concerning Competency-Based Testing..... 78
J. Stanley Ahmann (Iowa State University)

The Portland, Oregon Session

Competency Measurement at the Local Level: A Case Study of the Portland, Oregon Public Schools..... 91
Walter E. Hathaway

The Assessment of Competence: An Oregon Case Study.. 121
W. R. Nance

Competency Based Education..... 124
Gilbert Sax (University of Washington)

Discussion and Recommendations..... 133

NSPER: 78

Cincinnati, Ohio

CATHERINE CANDOR—CHANDLER
Kanawha County Schools, West Virginia

THOMAS H. FISHER
Florida State Department of Education

WILLIAM A. MEHRENS
Michigan State University

COMPETENCY MEASUREMENT AT THE LOCAL LEVEL: A CASE STUDY OF KANAWHA COUNTY SCHOOLS, WEST VIRGINIA

Catherine Candor-Chandler

Kanawha County, West Virginia, named for an Indian tribe that once lived within its borders, comprises an area of 914 square miles. A study in contrasts, the composition of Kanawha County is quite unique in its mixture of urban, suburban and rural areas as well as in the variations in socio-economic and educational levels of the population. The school system, the largest in the state, serves the entire county and has an enrollment of approximately 45,000 students with 40 percent from rural areas, 30 percent from the capitol city of Charleston and 30 percent from suburban communities.

Contrary to the national trend, standardized achievement test scores in the state of West Virginia and in Kanawha County have not been declining. In fact, an article in the *New York Times* in August, 1977, singled out West Virginia as a state that has not suffered a decline in test scores despite its relatively low expenditures for education.

Nationwide, most states or school systems are involved in the competency movement because of a legislative or state board of education mandate or public outcry over declining student achievement in basic skills. The impetus in Kanawha County came from neither of these sources but from a motion passed by the local board of education in February, 1976, asking the administration to study standards of achievement and/or multiple diplomas. Concern about grading and course requirements for special education students appears to have been the motivating factor behind this request, rather than a more general concern about the competencies of high school graduates.

THE PROCESS

As an initial phase of the response to the request from the board of education, a pilot study was completed in the spring of 1976 to assess the competencies of Kanawha County's high school graduating seniors. Criterion-referenced tests, measuring a preliminary list of proficiencies, were administered in reading and mathematics to sample populations of twelfth grade students in six of the county high schools. General trends identified in this study were:

1. Graduating seniors demonstrated acceptable performance in all areas of reading.

2. Proficiency in reading exceeded proficiency in mathematics.

3. As the complexity of basic computational skills increased, proficiency decreased. That is, while students could add, substract and multiply, many had difficulty with fractions, decimals and percent.

In the Fall of 1976 a steering committtee was named by Kanawha County Schools to begin formulating recommendations in response to the board's motion. The nine member steering committee included elementary, junior, and senior high school principals and central office representatives. After several meetings of the steering committee, it was decided that broader input was needed for the study.

In order to keep discussion groups at a manageable size and to maximize opportunities for input, three subscommittees were formed. Each subcommittee included about 20 persons and was chaired by a member of the steering committee. These subcommittees included: Exceptional Children, comprised of special educators, representatives from advocacy groups for specific exceptionalities and parents; Community Expectations, made up solely of major employers in the area; Subject Area, with representatives from students, teachers, curriculum specialists and local colleges.

Each subcommittee met at least four times to review options for improving student achievement in basic skill areas. The subcommittees were also asked to look at current graduation requirements in light of recommendations or suggested changes to make the diplomas more meaningful. Minimum competencies for high school graduation were explored in each subcommittee as one possible alternative approach. Subcommittee discussions were shared in the steering committee to see that the three groups were moving in similar directions. Admittedly, there were times when it appeared that it would be impossible to achieve consensus within a given subcommittee, let alone across the three subcommittees. The most united group was the exceptional children subcommittee.

The exceptional children subcommittee was adamant that any model considered by the board of education should not systematically:

1. deny exceptional children the opportunity to receive a high school diploma.

2. provide for differentiated diplomas, especially a "special education" diploma.

3. establish a different set of standards for exceptional children or for different types of exceptionalities.

4. waive requirements for exceptional children.

We finally decided to critically analyze all of the different models legislated or planned by various states or communties associated with

upgrading the worth of the high school diploma. The steering committee drew flow charts of each of the identified models (e.g., the Denver model, grade to grade promotion, differentiated diplomas). A brief description of the workings of each model as well as the graphic portrayal were presented to each subcommittee for review. Through these discussions, a list of constraints or restrictions on any high school graduation requirement model emerged. Concurrently, suggestions for positive change within the Kanawha County School system and identification of strengths in other models for improving student achievement in the basic skills occurred.

PRIMARY CONSIDERATIONS

Primary considerations generated by the subcommittees for the development of any model included:

1. Curriculum modifications should be minimal. The program should build on and improve the existing structure rather than require the installation of a new system.
2. The program design should require no drastic changes in staffing patterns.
3. The burden for implementing the program should not fall entirely on the classroom teacher, but should be shared by parents, students, teachers and administrators.
4. The model should be easily understood by the community.
5. The model should not require extensive additional testing, but should build on currently used diagnostic techniques.
6. The emphasis on revision of graduation requirements should not be on screening students out of a diploma but rather on identifying and remediating deficiences while at the same time verifying to the community that the end result of secondary education does meet the expectations of society.
7. Diagnosis, prescription and any necessary remediation must begin in the primary grades. The development of competencies in basic skills areas cannot wait until junior high and high school.
8. No group of students should be excluded from opportunities to meet graduation requirements.
9. The model must have a review procedure for clarification, appeal or waiver of competency requirements.

Additional recommendations were made for areas such as social promotion, special education, staffing, the relationship between academic skills and real life skills. Examples of these additional recommendations are:

1. Reading and basic arithmetic skills should be assessed initially with communications skills and other areas of concern to be considered later. A procedure for obtaining consensus on what are the minimum competencies in reading and mathematics needs to be established.

2. Carnegie unit and course requirements for graduation should be maintained to foster the student's incentive to achieve above the minimum.

3. A procedure to insure acceptance, understanding and a commitment to competencies among teachers and administrators and the community must be established.

It was from the exceptional children subcommittee that the suggestion first came to develop a range of skills in reading and mathematics to allow students from the more able to the less able to work toward a number of competencies. All graduates would receive a diploma and a competency certificate indicating skills the student had mastered.

After extensive debate and discussion in all three subcommittees and the steering committee, recommendations for a model to assess competencies were presented to the Kanawha County Board of Education. They had at their core the idea that all students who meet the Carnegie unit or credit requirement for high school graduation should receive a diploma and that all graduating seniors should receive the same kind of diploma. Certificates of competency were to be awarded to all students taking the competency tests, whether or not the diploma requirements were met. These certificates would indicate which of the specific groups of competencies were met, ranging from the minimal skills necessary to function in adult society that all persons (including the educable mentally retarded) should master to skills representative of a typical high school student's achievement.

The model unanimously adopted by the Kanawha County Board of Education in June, 1977, included competencies in reading and mathematics only, with expansion to other curriculum areas in the future.

1. When to Measure?

The model adopted in Kanawha County begins to assess competencies with informal checks of basic skills progress in the first grade and follows the student's progress until graduation. The purpose is to identify students with academic problems at an early stage so that such problems can be dealt with before they become more severe. Until the end of the third grade, these progress checks are to be carried out on a periodic basis by the classroom teacher. A K-3 program of studies detailing essential competencies and competency requirements has been developed to aid teachers in this task. Parents are to be notified about any deficiencies in reading or mathematics and school staff plans to help the student become proficient. In cases where conferences with parents are not possible, deficiencies should

be noted in writing. Communication with parents is a key component of his model at all levels.

In the third, sixth and ninth grades the Comprehensive Test of Basic Skills (CTBS) is administered to all students as a part of the West Virginia State-County Testing Program. As with most standardized achievement tests, the CTBS is a survey test designed for norm-referenced interpretations. It is, however, possible to use the results of the CTBS as a proficiency or criterion-referenced measure. Cutoff scores or criterion scores have been set for defining the competencies considered important by Kanawha County Schools. This process will be described in more detail later.

Since this process uses the existing CTBS data at grades three, six and nine to make criterion-referenced interpretations about student progress, no additional testing is required. The state-county testing data is reanalyzed as a global screening device to identify students with basic skill deficiencies. Students failing to meet the criterion level on the reading or mathematics subtests are identified by school. Parents are notified in writing of deficiencies found and the steps to be taken to correct the problem.

Basic competencies testing would begin at the tenth grade level with all students required to take these tests. Rather than develop an overly ambitious list of competencies, the focus is on the more fundamental skills associated with reading and mathematics. A student passing all sections of the tests would not take them again. Students failing the tests in either reading and/or mathematics would be provided with course offerings to improve their proficiencies in these basic skills. These students would be given at least yearly opportunities to pass these basic competency tests during high school. It would be mandatory for students who have not passed the basic competency tests to take them again during their senior year (Attachment 1).

These basic competency tests need to be paired with verification of formal school experiences to document a comprehensive education. Some purely cognitive products of experience may be readily measured, while others are difficult to evaluate. Falling in the latter category are things such as discussion, the response of others to one's actions, the completion of tasks, the assumption of responsibility and participation in group activities, all of which contribute significantly to a social and intellectual preparedness for adulthood. These benefits of actual classroom experiences can be central to education. Thus, credit and course requirements (Carnegie units) should be maintained not only to verify participation in certain classes, but also to motivate students to achieve beyong the minimum requirements.

Competency measures in reading and mathematics will be used to evaluate skill proficiency; credits will be used to document comple-

tion of courses and programs. Together they will make the evaluation picture complete. All students meeting the Carnegie unit requirements will be given a high school diploma. Certificates of competency will be awarded to all students whether or not the requirements of the diploma are met. The certificates will indicate which of the specified competencies have been attained. For the non-graduate these certificates could be seen as an interim record of progress. By receiving something upon leaving school, rather than nothing, students might be encouraged to return to complete their education.

The competency certificates will be attached to all diplomas and become a permanent part of a student's transcript. Employers or others reviewing a student's records would then be aware of whether or not that individual was proficient in the basic skills of reading and mathematics.

2. Who Selects The Competencies To Be Measured?

Our competency testing program has relied heavily on teacher input to select the competencies. Classroom teachers were responsible for developing the K-3 essential competency statements and minimum competency requirements to identify students with basic skills deficiencies. I want to emphasize that this is not grade to grade promotion. The K-3 program of studies is the basis for informal assessment of student progress on a regular basis by the classroom teacher. The results of these assessments will be used along with other information about a student (such as age, other achievement indicators, health) to make promotion/retention decisions.

The items that are used on the third, sixth and ninth grade CTBS tests were identified by classroom teachers. The process of item selection defines those competencies, measured by CTBS items, that teachers consider essential.

Likewise, the list of competencies for the high school program were written by teachers. One reading/language arts and one math teacher from each of our eleven high schools were selected from a pool of applicants to participate in this process. Curriculum specialists and others from the central office participated as resource people with the bulk of the writing, selecting and organizing carried out by the teachers.

One note of interest: curriculum guides or programs of studies were developed before developing the list of competencies. The competencies that were selected were taken from the curriculum. The curriculum of the school system shaped the competency tests rather than the tests determining the curriculum. This process virtually assured a match between the competency requirements and course offerings.

3. How To Measure?

I'd like to start the discussion of how to measure competencies with our program at grades three, six and nine before moving to the actual high school competency tests.

Traditionally, the responsibility for teaching the 3 R's has rested with the elementary schools. Competency testing programs beginning at the high school level shift the responsibility for demonstrating effective basic skills instruction from the elementary to the secondary schools.

Several states and school systems have recognized the need for early identification and remediation of basic skill deficiences. Measurement and instructional experts seem to agree that the use of criterion-referenced tests is more appropriate than norm-referenced tests for this process. Yet the costs for implementing such a criterion-referenced diagnostic testing system are prohibitively high. According to data prepared by Barry B. Anderson in, *The Competency Movement: Problems and Solutions*, the average cost per test item developed at the local level is about $100. The cost for just the administration and scoring of an existing criterion-referenced test available from a publishing firm ranges from about 50 cents to $2.00 per pupil.

Currently most school systems administer standardized achievement tests to their students on a regular basis. In an introduction to the *Myth of Measurability*, Paul L. Houts writes, "Last year alone, (sic 1976) over 40 million elementary and secondary school children underwent standardized testing at a cost of well over a quarter of a billion dollars."

The question we posed was whether existing standardized test data could provide a method for identifying student deficiencies prior to high school competency testing. In *Competency Tests and Graduation Requirements* the NASSP Task Force states that it is possible to use standardized achievement tests as performance or proficiency measures. "Cutoff scores or criterion levels can be set by a school district in defining the competencies it considers important." I will describe the results of our utilization of the data from an existing standardized testing program as a criterion-referenced diagnostic tool for elementary and junior high school students.

The Process

The Comprehensive Test of Basic Skills (CTBS) is a part of the West Virginia State-County Testing Program and is administered statewide to all students in grades three, six, nine and eleven. As with most standardized achievement tests, the CTBS is a survey test

designed for norm-referenced interpretations. Because one of the primary purposes of this type of test is to determine how a student or groups of students perform in comparison to a national group of students at the same level, the tests cover a broad range of skills.

The curricular content reflected in standardized tests is usually described as, "what is commonly taught in schools across the country." This determination is made most often by reviewing the textbooks of major publishers and conferring with subject matter specialists. In a typical school system there would not be a perfect match between the local curricula and the items on a standardized test. However, if the standardized achievement test has been carefully selected, the majority of the items on the test should assess "important" concepts of skills stressed in the school system.

The argument that a standardized achievement test contains many items that are appropriate for assessing a school system's curriculum is especially true for the basic skills areas of reading and mathematics. While the enabling or developmental skills at any level might vary, there certainly should be significant commonality between outcome measures of basic skills across the country if, indeed, they are "basic."

As you will recall, two of the primary considerations recommended by parents, educators and business leaders in the development of the competency plan were:

1. Diagnosis, prescription and any necessary remediation must begin in the primary grades. The development of competencies in basic skills areas cannot wait until junior high and high school.

2. The model should not require extensive additional testing, but should build on currently used diagnostic techniques.

Since the CTBS is administered yearly to all students in grades three, six and nine, it was decided that this data could be used as a global screening device to identify students with basic skills deficiences. There was strong sentiment from the committees that studied competencies that any such screening should be criterion-referenced rather than norm-referenced.

In order to make criterion-referenced interpretations of the CTBS data, teams of teachers representing a cross section of Kanawha County were brought in to review the content of the CTBS. Their task was to identify those items on the reading and mathematics sections of the test that measured essential competencies. For this purpose an essential competency was defined as a skill that a student should have mastered in order to advance to the next level without experiencing failure or frustration.

The teachers at each level, in both subject areas analyzed every CTBS test item related to the appropriate curriculum area. Item analysis data related to the performance of students in the national norming population, the state of West Virginia and Kanawha County as well as to the Kanawha County program of studies for reading

and mathematics and the currently adopted textbooks were used as references. Consensus had to be obtained on whether or not a particular item measured an essential competency. If consensus could not be achieved, it was decided that the item would not be included. A rationale for selecting or not selecting each item was written by the teachers. Typical comments for "rejected" items included: "Not a basic skill at this level;" "Too difficult;" "Irrevelant or impractical;" "Not stressed in the curriculum." Most of the selected items had *p* values or difficulty indices greater than .60.

A pool of 35 to 40 items for both reading and mathematics was selected for grades three, six and nine. These items then formed mini criterion-referenced tests. Although all essential competencies at each level are not tested, the content does reflect a significant sample of important skills.

Since students take the entire CTBS both norm-referenced and criterion-referenced interpretations are available from the same test. The norm-referenced interpretation is obtained through the usual process; the criterion-referenced interpretation is reached through manipulation of existing data. For the criterion-referenced application, the teacher-selected items are pulled from the computer tape. Those items are then reanalyzed as a criterion-referenced test.

GRADE 3.7

Total Reading

	Number of Items	*Reliability*	*Mean*	*SD*	*Mean Percent-of Correct Responses*
CTBS	85	.96	48.3	19.95	56.8
Mini CRT	40	.93	29.6	8.86	74.0
Total Math					
CTBS	98	.97	64.0	21.81	65.3
Mini CRT	36	.89	29.3	6.25	81.4

GRADE 6.7

Total Reading

CTBS	85	.96	57.9	18.75	68.1
Mini CRT	35	.90	28.9	6.18	82.5
Total Math					
CTBS	98	.96	68.4	19.85	69.8
Mini CRT	40	.90	33.8	6.72	84.5

The Results

The data presented above is from the spring of 1977. Two thousand, nine hundred and seventeen third level students and 3,330 sixth level students were tested. Data for 3,638 ninth level students tested in the fall of 1977 have been analyzed with very similar results.

As can be seen from the tables above, the mini criterion-referenced tests are much "easier" than the total CTBS. The mean percentage of correct responses varies from 56.8 to 69.8 on the CTBS and from 74.0 to 84.5 percent on the mini criterion-referenced tests. This was accomplished without any significant decline in reliability (KR-21) coefficients.

The purpose of such mini CRT's is to identify students with reading and/or mathematics skill deficiencies. Since the test is easy, these scores discriminate better among those with low achievement than the CTBS. Both the percentage of students scoring below the chance level and the percentage of students with less than 50 percent correct responses are reduced. The following table summarizes this data.

	Percent of students scoring below chance level	*Percent of students with less than 50% correct responses*
GRADE 3.7		
Reading		
CTBS	8	41
Mini CRT	2	18
Mathematics		
CTBS	5	27
Mini CRT	less than 1	7
GRADE 6.7		
Reading		
CTBS	3	22
Mini CRT	less than 1	8
Mathematics		
CTBS	2	21
Mini CRT	less than 1	6

By making the test easier, the reliability of the results increased for low achieving students and decreased for high achieving students. In setting a criterion-level to identify those students with basic skills deficiencies, the probability of making a false negative error is reduced.

This process is viewed as a more valid and reliable approach to the global diagnostic screening of students for reading and mathematics skill deficiencies than using standardized achievement data provided by test publishers. The item selection process assures congruency between items and essential competencies in a local curriculum. Also, a criterion-referenced approach to student identification overcomes many of the disadvantages of norm-referenced scores. Two major advantages of these mini CRT's are that they more nearly reflect specific instructional purposes and they evaluate how close students are to achieving mastery of objectives rather than where they stand in relation to some average or norm. Further diagnostic testing would still be necessary to isolate the specific skill weaknesses of a student with a reading or mathematics deficiency. Rather than eliminating the need for careful diagnosis, this process provides a global screening device to identify those who are in need of such diagnostic testing. The number of students to be tested is reduced, not the need for a good diagnostic test.

The high school competency tests are different. Criterion-referenced tests are currently being developed and field tested. No competency tests have actually been produced yet, although sets of items have been tested. Fortunately when our plan was adopted by the board, we were given two and one-half years of test development. Our timeline looks like this:

Schedule for Developing Kanawha County Minimum Competency Tests

Dates	*Activities*
Completed	Develop objectives.
Completed	Develop program plan.
February 1978	Test specifications and item writing workshop. Additional item writing activity as needed.
March-April 1978	Item editing and review. Production of pretests.
May 1978	Pretesting of questions for "pool."
June 1978	Analysis of pretests.
July-August 1978	Review of pretest results. Review of status of pool of questions. Additional item writing and review.
September-October 1978	Item editing and review. Production of pretests #2.

November- December 1978	Pretest #2 administered.
January 1979	Analysis of pretests #2.
February 1979	Review pretest results.
March-April 1979	Develop final tests for operational year one.
May-June 1979	Final test assembly, year one tests. Final committee review.
July-August 1979	Printing, distribution of final tests.
September 1979	First use.

All tests are paper and pencil, multiple choice tests. This decision was made because the state of the art in measurement appeared to be more advanced for this type of testing than in other types such as performance tests. Recognizing the limitations of paper and pencil tests and because the results of the basic competencies test will become a permanent part of the students' records, a review procedure for clarification or appeal of the results of such tests is a part of the competency plan. In exceptional situations alternate methods of verification of competencies or a waiver of these requirements might be necessary. An example of the former category, that of alternative methods of verification, could be a physically handicapped student unable to complete traditional paper and pencil tests, but possessing all of the necessary basic skills. Another example could be a student with test phobia. A case where a waiver might be considered would be that of a non-English speaking student moving with his family to Kanawha County from Sweden during his senior year. However, procedures for this review process must be carefully delineated to maintain the integrity of the competencies certificates.

4. Who Determines How To Measure Them?

Once the competencies were identified for the high school competency tests, test specifications were developed locally by myself and a couple of other central office staff members. Dr. John Fremer, from Education Testing Service, worked with us as a consultant to review our plans for technical adequacy and to train teachers to write the actual test items.

In terms of who determines how to measure the competencies, the decision to use paper and pencil multiple choice tests was made basically by me, because they are the least expensive and easiest to develop, administer and score. Different kinds of tests may give different results. However, we are assessing in school achievement and, paper and pencil tests are the most commonly used for this purpose. So our method of assessing is typical of other assessments during a student's academic career.

Also, remember we have provisions for verification, appeal or waiver and have accepted the responsibility of designing alternative means of assessment for those who feel that paper and pencil tests do not document their mastery of the basic skills.

5. What Competencies?

Since all students, including exceptional children, must take the high school competency tests, a broader range of competencies were included for assessment than on most competency tests. Before I describe the high school competency tests, I'd like to describe our rationale for selecting the competencies.

In many quarters there is considerable debate over whether to assess school skills or life skills. Again, let me remind you that one of the primary considerations in the development of our model was that curriculum modifications should be minimal, building on and improving the existing structure rather than requiring the installation of a new system.

Our choice was to measure school skills since that's what our curriculum is based on. It seems obvious that most "school skills" in a given curriculum must have been selected because of some relationship to those more terminal abilities we call life skills. There are, however, many life skills that cannot be found in the school curriculum. Rather than restructure the school experience to include opportunities to develop competencies in a vast array of life skills, we used the school skill base and selected those competencies which also have applications as life skills. The ability to balance a checkbook or read a newspaper or figure out how much paint will cover a wall or complete a job application are examples of life skills included in our list of competencies that we feel have their basis in school skills. The distinction between school skills and life skills becomes increasingly indistinct using this logic. Let me go through our high school competency test with you now.

The mathematics portion of the test currently has six subtests. They are: addition and subtraction of whole numbers; multiplication and division of whole numbers; fractions; decimals and percent. These four are entirely non-verbal. No reading whatsoever is required to do these problems. Probably these tests would be categorized more as school skills than life skills. Although if these school skills have not been developed, it would be highly improbable that the concomitant life skills have either. The other two mathematics subtests are measurement and informal geometry and consumer applications. The measurement and informal geometry test does not require knowledge of high school geometry but rather focuses on the application of those measurement and geometry skills that permeate the curriculum beginning in elementary school. Examples would be

time, maps and charts, distance, weight, and area. All of these skills, I think, could be classified as life skills. That statement is also true for the consumer application section. This test requires students to apply the four basic mathematical operations in real life situations. Examples are unit pricing, the commission on a sale, and determining wages.

The reading portion has four subtests. The first is functional literacy. This parallels other functional literacy tests that I'm sure you have seen and assesses those minimal reading skills necessary to function as an adult in society. We hope almost all of our EMR students will be able to attain these competencies. Next comes functional vocabulary, a step above functional literacy. This subtest checks a student's mastery of that vocabulary encountered in the real world from fields such as education, law, medicine and business. The third section is mass communication skills. This section assesses a student's competency in dealing with printed mass media such as the newspaper and magazines as well as skills such as interpreting ads and ordering from a catalogue. The last subtest is labeled reading comprehension and is perhaps most typical of a regular reading achievement test at first glance. The test measures a student's ability to read and comprehend selections that are typical of those things someone might read for enjoyment as an adult. The determination of whether this is a school skill or a life skill is hazy. Selections have been chosen from "great" literature, poetry, plays and essays as well as from such other sources, such as social commentaries and scientific articles.

The organization and structure of our tests leads directly to the next question of one minimum or many.

6. One Minimum or Many?

Remember that in our early planning discussions of a model to upgrade the worth of a high school diploma, the exceptional children subcommittee was adamant that different sets of standards should not be established for exceptional children or for different types of exceptionalities. Thus, the minimum passing score will be the same for all students.

The ten subtests just described, six in math and four in reading, range in difficulty from minimum to typical. By that I mean that the easiest of the subtests should be mastered by all or nearly all students while the most challenging should be mastered by the typical or average high school student. None of the tests would be a challenge, hopefully, to top students. We anticipate that up to 20 percent of our student population may not attain the most difficult competencies.

We have one minimum for each test, but many tests. The differential expectations are handled not through different standards

but rather through the assessment of more complex competencies. Since high school graduation is not dependent on passing any of these tests, we feel comfortable taking this approach. The emphasis is a positive one on the documentation of competencies attained rather than on what a student has failed. Even students with limited ability should have documentation on their transcript that they do, in fact, possess certain basic skills, while an average high school student would have documentation that he/she possesses skills beyond "functional literacy" assessed in most competency testing programs.

7. Who Sets The Minimums?

Perhaps a better question would be how to set the minimums since the who should be answered as a part of the how. I wish I had some successful experiences to report. I will describe two approaches tried so far that have had limited success.

First we had the reading and mathematics curriculum specialists identify performance statements to differentiate competent from incompetent seventh grade students. These lists were sent to the seventh grade teachers. They were asked to check for each student the competencies that were not demonstrated by the student in their judgment through observation of classroom performance. The intent was to compare teacher assessment of student competence with tests scores, in this case the score on the sixth grade CTBS mini CRT. We discovered two things in this process:

1. Some teachers had no idea whether or not students were competent. Some of the responses were rather frightening. For example one of the performance statements in math was, "The student understands the concept of fractional parts." One teacher said, "How would I know. All of my students are working in the textbook on whole number operations and we haven't gotten there yet." One of the reading performance indicators was that the student could read a simple paragraph with expression. A teacher commented, "How could I tell that, I don't even know if he can talk."

2. The correlation between test scores and teacher judgment varied considerably among schools. The correlation coefficient was as low as .30 in one case. County-wide the best correlation between teacher judgment of competency and any minimum standard was about .60.

This approach was abandoned because we felt that we needed to plan additional training with teachers about what we meant by competency if we were to achieve any uniformity of results.

The second approach, attempted once again on the mini CRTs generated from the CTBS, was taken from the ETS manual on setting standards by Zeiky and Livingston. The approach we selected, which can only be used on multiple choice items, was Nedelsky's

method. The instructions for this approach are as follows:

1. Make a copy of the Judge's Recording Form: Nedelsky Method for each judge (Attachment 2).

2. Give each judge a copy of the test and the recording form. Have the judges look at the first question and identify the wrong answers that a minimally competent student would be able to recognize as wrong. For example, consider the following question. It is based on a letter from a job applicant to an employer. The student's task is to choose the word or phrase that best completes the sentence:

"I would like to be considered for the position of secretary, _____ I do not have any experience in typing."

A) but B) and C) even though D) because

A judge might decide that a minimally competent student should be able to eliminate wrong answer D because it makes the sentence absurd. The judge might also expect the minimally competent student to recognize that wrong answer B is not as good as A or C, since it does not indicate that the writer realizes that experience is desirable. But the judge might decide that the choice between wrong answer A and the correct answer C is too subtle for the minimally competent student. The judge would then identify answers B and D as being so clearly wrong that the minimally competent student should be able to recognize that those answers are wrong.

3. Conduct a brief discussion of each test question, using the following format:

A. Focus the judges' attention on the first wrong answer. Ask how many thought the minimally competent student should be able to eliminate it and how many did not think so.

B. If the judges are not unanimous, ask for one judge who identified that wrong answer to explain why. Then ask for one judge who did not identify it to explain why not. Do not try to reach agreement; just allow each point of view to be heard. The judges may or may not be swayed by the comments of their colleagues.

C. Go on to the next wrong answer.

4. After the question has been discussed in this way, ask the judges to circle on their recording forms the number of wrong answers identified. Then go on to the next question. Continue until all the questions in the text have been completed. Then collect the forms from the judges.

To combine the judgments to set a standard, follow the steps outlined in the next column.

1. Convert each Judge's Recording Form to a table of probabilities. Conceptually, the way to do this is to assume that the student will guess at random from among the answers he or she does not know to be wrong. For example, if a question has four possible

answers and the student can rule out one of them, the student will have three answers left to choose from. This gives the student a one-third chance of guessing correctly. Operationally, the process is a simple clerical one. Take each Judge's Recording Form. For each question, in the column headed P, write:

".25" (if the judge circled 0)
".33" (if the judge circled 1)
".50" (if the judge circled 2)
"1.00" (if the judge cirlced 3)

2. Use a hand calculator to add up the numbers in column P on each form. Write the sume in the box labeled SUM.

3. Now just calculate the verage of the sums. Simply add all the sums and divide by the number of sums. The number of sums should, of course, be equal to the number of judges.

The result is an estimate of the standard for the test.[1]

Although this process produced relatively uniform standards across groups of judges, we were not satisfied with the approach for two reasons: 1) Teachers (who were our judges in the three pilot attempts at this approach) tended to estimate the probablity of students getting "easier" items right lower than the "harder" items. In discussions afterwards, it seems this was justified as follows. "For easy questions students could make a simple error, but for more difficult ones they either know it or don't know it." The definition of what is meant by "minimally competent" is not clear. Each individual seems to have their own definition. 2) Teachers felt very uncomfortable with the process. Comments such as "I'm just guessing; or how would I know" were frequently heard. Thus, the teachers did not feel good about the standards set because they didn't feel good about the process used to obtain it. The consistency of the results across groups was not enough to overcome the feelings of arbitrariness about the process.

Consequently, the approach used to set the standard for the third, sixth and ninth grade tests was totally arbitrary. After looking at preliminary data and using a touch of common sense, I picked 60 percent as the criterion. This decision can be rationalized to some degree since these results are used only for screening, i.e., to identify students who *may* have basic skills deficiences.

The standards for the high school competency tests have not yet been established. Hopefully by next fall, in time for the first administration of the tests, a satisfactory standard setting procedure will have been identified. Who sets the minimums will be dependent upon the procedure used. Hopefully a cross section of people (teachers, parents, students, "successful" adults) would be used.

8. How High The Minimum?

Now we are getting to the easier questions. How high the minimum is not a real concern for us because the competency tests are independent of the requirements for high school graduation. The series of mathematics and reading subtests reflect the philosophy of "each student to his/her own potential." Some of the subtests may be too easy for above average students and some may be too hard for below average students. The range of skills covered in the competency subtests should be sufficient to provide some reachable goals for the less able students and an interesting exercise for the more able students.

9. Minimums for Students or Schools?

The standards set are for each student. Test results are reported to individual parents with summary results reported to the general public. As with other types of test results, it is likely that the competency test results in the public's mind will also be used to set minimum for schools. It seems to be human nature to want to know who did the best, who did the worst and where one stands comparatively.

By not tying the competency test results to high school graduation, the minimums set for students are less punitive than other types of programs. We do not want students to have to take the whole rap for "incompetence" but rather we want to analyze our school programs in a process of improvement.

10. What to do With the Incompetents?

First of all, school students will be given at least yearly opportunities to pass the competency tests. All students will take the reading and mathematics tests in the tenth grade. A student passing all sections of the tests would not take the tests again. Students failing any section of the tests in either mathematics or reading would be provided with course offerings to improve their mastery of these basic skills.

There are three things I would like to point out. First, the basic skills courses in reading and mathematics at the high school level will probably be electives. We expect the student to show initiative in terms of selecting those courses, with the help of his/her parents, teachers and counselor, that will maximize basic skills achievement. We feel strongly that the responsibility for learning must be shared. The schools cannot do it all by themselves. Secondly, with screening beginning at the first and second grade levels, we hope to emphasize the basic skills at the primary level. Basic skills have traditionally

been taught and mastered in elementary schools. We want the junior and senior high school programs to have to emphasize the basics for only those few students who have yet to master them. We hope the secondary curriculum can retain its diversity and vitality and continue to meet the needs of the majority of students who have already mastered the basics.

Thirdly, we have a new program in operation through which funds are provided to each secondary school to design its own plans to help the "incompetents." Our schools are different with each having its own set of unique characteristics. In their planning, schools may consider any alternative including such things as inservice training for teachers, parent workshops or programs to improve school attendance.

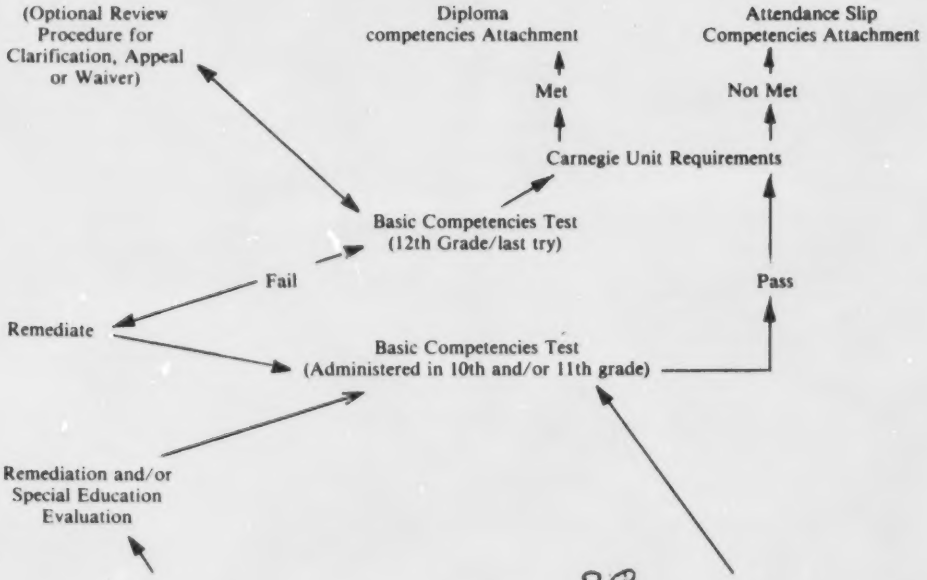
Our summer school program is yet another potential solution to increase basic skills achievement. We also have a new volunteer tutor program designed to provide individual attention from trained community personnel to students desiring help.

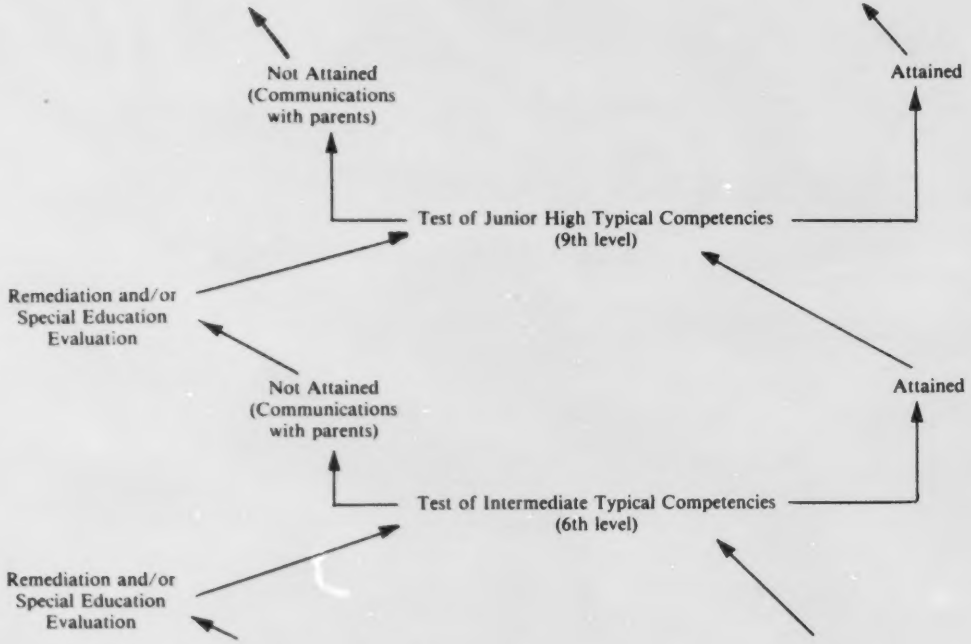
We do not think we can solve all the problems associated with education or make all students basic skills whizzes. However, we do see this as an opportunity to give increased emphasis to the pursuit of academic excellence in all curricular programs. Hopefully, these efforts and the involvement and commitment of the entire school community will serve as a vehicle for continued confidence in the school system.

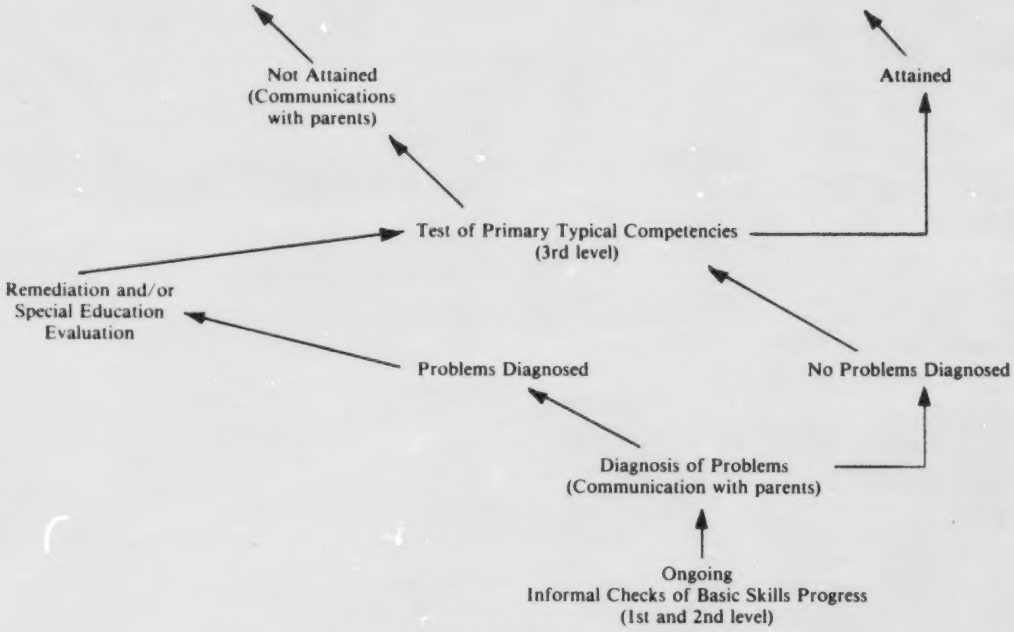
Notes

From Michael J. Zieky and Samuel A. Livingston, *Manual for Setting Standards on the Basic Skills Assessment Tests.* Copyright © 1977 by Educational Testing Service. All rights reserved. Reprinted by permission.

Attachment 1




Not Attained
(Communications
with parents)
Attained
Test of Junior High Typical Competencies
(9th level)
Remediation and/or
Special Education
Evaluation
Not Attained
(Communications
with parents)
Attained
Test of Intermediate Typical Competencies
(6th level)
Remediation and/or
Special Education
Evaluation


Not Attained
(Communications
with parents)
Attained
Test of Primary Typical Competencies
(3rd level)
Remediation and/or
Special Education
Evaluation
Problems Diagnosed
No Problems Diagnosed
Diagnosis of Problems
(Communication with parents)
Ongoing
Informal Checks of Basic Skills Progress
(1st and 2nd level)

Attachment 2

Judge's Recording Form
Nedelsky Method

Test Title: _____

Judge's Name: _____

Question Number	*Circle Number of Choices identified*				*P*
1	0	1	2	3	
2	0	1	2	3	
3	0	1	2	3	
4	0	1	2	3	
5	0	1	2	3	
6	0	1	2	3	
7	0	1	2	3	
8	0	1	2	3	
9	0	1	2	3	
10	0	1	2	3	
				Sum	

THE ASSESSMENT OF STUDENT COMPETENCIES IN FLORIDA'S PUBLIC SCHOOLS

Thomas H. Fisher

Introduction

In a recent presentation to the prestigious Florida Education Council, Carey Ferrell, Director of Public Schools for the State of Florida, stated that the new competency test for high school juniors had created more positive changes in Florida's public schools than anything he had seen in his seventeen years of service in public education. At the opposite end of the enthisiasm continuum, attorneys for Florida Legal Services completed on August 23, 1978, a legal challenge against the scoring system used in the competency tests. This case was largely based on the assertion that the tests are unconstitutional, vague, and arbitrary.

It is highly likely that other states with competency programs will face similar extremes in opinions. They should be able to profit from the experience of those states which were first to develop competency tests. The legal arguments made in the Florida situation suggest areas in the development of these tests that are particularly sensitive and vulnerable to attack.

The purpose of this paper is to describe how the State of Florida proceeded in its development of a competency testing program. Suggestions will be made to assist the reader in avoiding pit-falls wherever possible. Of course not all litigation has been completed in Florida; indeed, only the opening skirmishes have been seen. The reader will have to "stay tuned for further action."

Defining Minimum Competencies

1. Responsibility for Definitions

The general history of the development of the minimal competency testing program in Florida was clearly delineated in the May, 1978 issue of the *Phi Delta Kappan* and will not be repeated herein. Generally, the article showed that the Florida program evolved over a period of years. It is an extension of a growing concern among parents and citizens about the levels of achievement of students at all stages of education. It represents an attempt to look at whether or not students are learning.

Throughout the development of the various accountability laws in Florida, a central interest was expressed in the articulation of specific learner objectives. The Department of Education led the state in the creation of sets of objectives in several content areas in the early seventies. Generally, these objectives were written in highly detailed formats. The objectives were also rather specific in their content which made them difficult to implement in the schools.

From a process standpoint, the department was aware, even in those days, of the importance of "ownership" of the objectives. The objectives were derived from sets of existing objectives wherever possible, but the department was careful to make certain that appropriate groups of educators and influential people around the state had input into the process and approved the final products. This pattern was maintained and is used today by the department.

The early experience with highly detailed objectives led to frustration by teachers and a certain amount of ill-will toward the term "behavioral objective" itself. As the statewide assessment proceeded to measure these objectives, the concern mounted and it became evident that the department should move in the direction of making objectives more general.

By 1976, the Department of Education had developed a set of minimal objectives for grades 3, 5, 8 and 11 in the areas of reading, writing, and arithmetic. These objectives were broadly grouped into "milestone" objectives thus providing a great reduction in quantity. Two levels of objectives were developed for ease in communicating to teachers, parents, and students. In 1977, these objectives were transformed to the present set of minimum student performance standards and skills that form the basis of the present competency tests.

Again, the process used by the Department of Education stressed that the responsibility for the creation of the minimal objectives was the department's, but the input of local district educators was highly important. Typically, the Department of Education curriculum specialists gathered together a small group of subject area specialists who drafted the minimal objectives. The objectives were then sent to each school district along with a survey form that local educators could use to summarize their suggestions for changes. Meetings were then held across the state to gather additional suggestions relative to the revised objectives. Department curriculum specialists again revised the lists and the State Board of Education approved them. The entire process took about one year.

The Florida experience to date leads to the following suggestions for those building their sets of minimal objectives:

1. Allow enough time for the process. One year is an acceptable minimum.

2. Utilize local district educators to react to draft sets of objectives. Document the extent of involvement and agreement.

3. Utilize state curriculum organizations to assist in both the original drafting of objectives and the later review. It would not be inappropriate to contract with a state organization for the creation of objectives.

4. In the beginning, determine whether your objectives will represent things students are expected to know now or things you want them to know in the future.

5. Keep your objectives simple and brief to ease communication and utilization. Do not go overboard and write too many objectives.

6. Look ahead to the measurement issues and obtain technical assistance at an early date. Once the objectives are adopted it is almost impossible to make changes if it is discovered that an objective is not clearly worded.

7. Provide for a cyclical renewal of the objectives, perhaps on a three-year basis. Seeing the test data always causes people to change their minds about their expectations.

8. Make sure you have enough money to print copies of the objectives and send them to every school.

9. Don't expect to get everyone to agree. They won't.

10. Document everything and summarize the process in a brief report. Your attorneys may need it for an exhibit.

2. Scope of Competencies

The citizens of Florida have been consistent in their demands that the accountability program primarily focus on the basic skills of reading, writing and arithmetic. A few years ago there was a science assessment but it was dropped and there has not been any great impetus to revive it.

Other states have adopted other postures on this issue. Pennsylvania, for example, was one of the first states to have a statewide assessment program. It attempted to measure ten basic educational goals including various attitudinal dimensions.

Part of the problem in defining areas in which objectives should be created is the fact that most state education agencies have created sets of educational goals that are very broad. Some of these statements contain the proverbial kitchen sink because they are usually developed in a context that encourages the inclusion of goals dear to the hearts of certain influential people or organizations. These statements often contain mixtures of process and product goals and statements of philosophy as if they were all the same. Additionally, all curriculum groups want equal treatment in such statements of goals, so little attempt is made to differentiate among them. An exception to this can be seen in some of the early work in

"needs assessments" in certain states that led to prioritization of goal statements.

There can be no denying that the attention getting nature of a competency based program will create instant importance for whatever is measured. There will be great pressure to measure everything to insure that no areas are overlooked. In state or district programs that utilize a sampling approach, this probably will not be a problem, since many areas can be efficiently measured. But, in any student centered competency program, extreme breadth can mean the test *becomes* the curriculum. There won't be time for anything else.

Secondly, the Florida experience indicates that is is not possible to assemble even an elementary mathematics test with complete unanimity of curriculum and testing experts. Yet, mathematics is perhaps one of the most clearly definable curriculum areas. If agreement cannot be reached on measurement strategies for mathematics, then how can one ever expect to reach agreement in the social studies or, worse, in affective areas?

From a legal point of view, one can expect the plaintiffs to argue that not all children have had a proper opportunity to learn the required skills or that the classrooms and schools across the state are not equally teaching the skills. This only compounds the problems mentioned above. Implementors of competency testing programs have to agree on the objectives and on the "correct" responses desired from the students.

The Florida experience leads to the following suggestions:

1. Concentrate on the three R's of reading, writing, and arithmetic. Do not make life hard for yourself by biting off more than you can chew.
2. If your state is interested in areas outside the three R's concentrate on a cyclical sampling approach to gather data on a statewide basis.
3. Stay away from attitudinal assessments. Your opponents can really tear you apart on this one.
4. Limit your lists of objectives to school skills. At least there is general agreement on who is responsible for teaching them.

Similarly, you should seriously consider staying out of the "life skills" arena. School skills are a lot safer and easier to defend. This contrast in approaches can best be illustrated by the approach taken by the Palm Beach School Board compared to the State of Florida approach. Palm Beach adopted the Adult Proficiency Level test and suggested it truly measured those things necessary to survive in society. The state stuck with the three R's and made no specific claims that success on the test brings success in life. In retrospect, the state will probably be in a better position to defend its position that will Palm Beach.

3. Responsibility for Measurement

There is no easy guideline for determining whether the local or state agency is responsible for the measurement of minimal competencies. States differ widely in the strength of their local districts and state agencies. The people in one state can be expected to have a different philosophy of government than in another state. Some states expect and accept a strong central control whereas others demand maximum local control. What works in Florida may not work elsewhere, and it would be presumptuous to suggest that it would.

In Florida, the legislature has maintained a strong interest in educational accountability. The state agency was placed in a situation that demanded a great amount of centralization for implementing the educational assessment program. From a philosophical point of view, the use of a strong state approach can be justified in the interests of equal educational opportunities. A student in Sopchoppy, Florida, in the northern panhandle, should be provided the same opportunities to learn to read as a student in Miami. The Florida system is built around common data, the identification of relative needs, and the provision of state compensatory aid monies to assist low scoring districts. This philosophy cannot be implemented with a system that permits each district "to do its own thing."

From a legal point of view, some interesting arguments could be made by students in one district or school who were deprived of a diploma while students in an adjoining school district were permitted to graduate because standards were lower. This kind of discrepancy has been tolerated in the past because schools weren't competency based; it may not be tolerated in the future.

4. Timing of the Measurements

There were legislators in Flordia who thought in terms of a high school exit test that would be given in May of the student's senior year. The student would truly either jump the hurdle or not, with no chance for corrective action. To be more equitable, the decision was made to make the graduation test part of the grade eleven statewide assessment, which is administered in October.

This brings to mind a first important point; competency tests should be given early enough to permit corrective actions to take place. Students do not particularly mind new graduation requirements if they have a chance to learn things they have forgotten or did not quite master.

In some states, it has been proposed that the competency test be given as early as the eighth or ninth grade. If a student passes, he or

she is "competent" and does not have to worry about it again. Unfortunately, one never would know how much of that knowledge the student has forgotten by the time graduation rolls around. It might be argued that the school's responsibility is to send young adults into the world with the proper skills; it does not help to send them out with the confidence that they had the skills in eighth or ninth grade. Employers will not appreciate that kind of statement from the schools.

A second major consideration is the amount of lead-time to be given students who are facing competency requirements for the first time. This may mean the first time the test is implemented or it may mean the first time a modification in a standing test is introduced. This is an important consideration for districts and states currently preparing to administer tests for the first time.

In Florida, the 1976 Accountability Act went into effect July 1, 1976. For the first year, many districts and schools did little, believing that the requirement would be repealed by the 1977 Legislature. In October, 1977, the test was given for the first time to high school juniors and large numbers of them failed. The realization began to hit local district educators that something indeed was happening.

About the time the results began to sink in, the lament was heard. "It is unfair to start with the current group of juniors. They didn't have enough lead-time. You should have given at least four years lead-time." Many Florida students were interviewed by the newspapers after the results were made public, and they did not seem to have any complaints other than, 'Why me?'

Given the context of the situation, this reaction is probably going to be seen in other states as well. There probably are those who would argue that no amount of lead-time is fair until you can prove that all the students have had an equal chance to master the skills and actually have mastered them. In that case, one hardly would need the test.

Realistically, a start has to be made somewhere, and some amount of lead-time is necessary to develop the strategies and materials needed to give adequate warning to the students. It is not known at this time whether or not the two year lead-time given in Florida will be defended successfully in court. It is also not known exactly how many students, in the end, actually will not graduate. This would have an effect on whether or not one could accuse the program of being unfair.

In retrospect, it seems best to take as much lead-time as can be allotted, while remembering that momentum must be maintained to cause change. If you opt for long lead-times, then try to determine how you can "keep up a good head of steam." If local educators ever think you are not serious, nothing will happen in the local schools. In that case, the plaintiffs could conceivably bring suit

against the local school authorities for not following the leadership or dictates of the state authorities and the legislature.

5. Selection of Skills To Be Measured

Even if the minimal expectations are restricted to the areas of reading, writing and mathematics there will still remain the dilemma of which skills to measure. The educator is caught in a no-win trade-off between test length, which has a bearing on accuracy of classification to mastery status, and scope of the test. Will you measure many things with less accuracy or relatively few things with greater accuracy?

Beyond the aforementioned question, a second appears. Namely, should one articulate objectives that are minimal but do not lend themselves to paper and pencil tests? In other words, should one capitalize on the visibility of minimal objectives, be they state or local, to get teachers and students focused on certain skills for instructional reasons?

In Florida, the second question was answered affirmatively. The set of statewide minimal standards and skills included some that could be best described as performance oriented. For example, third graders were expected to dictate a sentence correctly. This was thought to be an important pre-writing skill and was included because teachers needed to focus attention on it if later objectives were to be measured. The Statewide Assessment Program's position was that it would measure as many of the minimal standards as could be defensibly included. Others would either not be measured at all or would be measured only on a statewide basis.

This gives a hint about the answer to the first question. Do not be too concerned if it is impossible to measure every skill. In the first place, there is nothing to prevent local district educators from gathering additional information to indicate student mastery on any of the basic skills if they so desire. Secondly, there will always be a limit on the length of any state or local competency test, and one should not be frustrated by that fact. Determine the highest priorities among the minimal skills that can be measured and measure them.

This brings up a related point. If the minimal standards and skills are too precise, too specific, too minute, measurement will be a nightmare. It would be better to make certain that the more general and broad level standards are included rather than get bogged down in minutiae.

6. One Minimum vs. Many

Other authors have from time to time written that it is somehow unfair to set one standard for all children and that it would be better to set variable standards for children based on their ability, family background, or other factors. The idea evidently is to make the process somehow more democratic by letting each person proceed at his or her own rate.

While this sounds like a laudable way to proceed, it also could be completely unworkable. In the first place, it is doubtful that educators have any way to determine what a person is ultimately capable of learning or with what speed. Secondly, in the basic skills of reading, writing, and mathematics it does not seem to be in the best interest of society to say that minimums cannot be set for at least the vast majority of students. (Note that this does not imply that trainable mentally retarded students, for example, must have the same required minimum standards. Indeed, in Florida they have their own standards.)

Another frequently heard argument is that smart students will not be challenged by minimum standards and their education will suffer. On the contrary, even smart students should master the basic skills and, once they have done this, it would be a sad reflection on their teachers if they were not encouraged to go beyond the minimum. In Florida no evidence has been forthcoming to demonstrate that the minimums have become the maximums.

7. What Constitutes A Minimum Competence?

A better way to phrase that point would be, "How high should the minimum be?" Or, "How many students can your state or district afford to fail?" This may be a backward way to begin, but, in a sense, it represents the political realities. Adjustments will be made throughout the educational system as standards are adopted, tests are administered, and remediation programs either succeed or do not.

But, clearly, there is also something dishonest about such questions. They seem to say that there is no present dissatisfaction with student achievements. If too many people seem to be flunking, the passing scores or the expectations can be changed. In other words, do not worry about the quality of instruction, the participation of parents in the schooling process, the quality of administrative leadership, etc. If the evidence is bad, bury it. Change the test.

It would be intellectually cleaner if educators and citizens were willing to set their expectations, agree on a test or other evaluation methodology, and then let the chips fall where they may. If too many students fail, recheck the standards and the tests, of course, but do

not simply take the easy way out and adjust the standards until the "right" number pass.

The approach taken in Florida has involved the adoption of minimum standards only after a year long developmental process. Then, the standards were adopted for a three year period. During the three years, data will be collected and the opinions of experts will be sought. Changes may be made in either direction, harder or easier, but only after serious consideration is given to the recommended changes.

When the 1977 functional literacy test results were released in Florida, the statewide data revealed eight percent failed in communications and 36 percent failed in mathematics. Some educators immediately stated that the math test was too hard, and the reading test was "just right." It did not take much to understand why these opinions were voiced. In reality, only one thing can be said about the results. Compared to the expectations, students did better in reading than they did in mathematics. A second question is whether or not the expectations were too low or too high.

One last thought on this matter, if you are establishing minimal expectations for graduation from high school, the hard realities are that a sixth or seventh grade reading level and an eighth or ninth grade math ability are the ball parks in which you will most likely play. Some citizens are going to be very angry that more cannot be done right away. Realistically, you will be lucky to achieve that goal in the next five years.

A Summary of Recommendations Regarding Measurement

The measurement issues relative to competency programs have only briefly and generally been addressed in the previous pages. Many more psychometric problems were not mentioned and go beyond the scope of this paper. It is still useful to summarize the most salient suggestions derived from the preceeding discussion based, again, on experiences in Florida.

1. Each state must determine for itself whether or not there will be a centralized approach to competency programs or if the responsibility will be given to the local agencies.
2. Do not be afraid of state minimum standards. They have the advantage of furthering equality of opportunity across a state and concentrating a state's resources.
3. Provide adequate phase-in time for the competency program.
4. Do not impose the competency test at too high or too low a grade level. Grades ten or eleven would be best.
5. Once the commitment is made to have a competency program, do not back down. Keep up the pressure and leadership.

6. Do not be afraid to use paper and pencil tests or to require additional evaluative information from local teachers.

7. Measure only those skills that are most important and have the most defensible measurement strategies.

8. Do not be afraid to use a common set of standards for all students across a state or district. Improved education will result if educators can at least accomplish that much.

9. Expect the entire system to recycle itself, but try to be objective. Do not "adjust" the system just to make the school look either good or bad.

8. Competent Students and Competent Schools

One of the most common problems created by a local or statewide assessment program is that of proper interpretation of the accumulated data. Test programs generate data that will be of interest to parents and citizens. Newspapers will give prominent coverage to reports of school and district test scores, much to the chagrin of the teachers and administrators. Publication of the data does not cause problem; interpretations cause the problem.

Simply stated, the problem is caused by the possibility that two different types of conclusions may be drawn from a given set of school or district test data. Low scores for a particular school may be a proper reflection of the current status of student achievement, but the scores may not be an adequate reflection of the efficiency of the school. The same set of data may be used to determine status and to evaluate the efficiency of the school, depending on who reads the results.

In terms of minimum competencies, it is clear that current programs seem to emphasize that the competencies are for the students. Yet, when the data are accumulated into school averages they become minimums for the school. For example, Johnny B. may be expected to add any two digit numbers as a fifth grade student. When the school reports the average number of students who have attained this objective as, perhaps 45 percent, the following interpretations can be expected:

1. The students are doing just fine; 45 percent is a large number to have mastered this skill.

2. The students in this school have a serious educational need because this important minimal objective is only being mastered by 45 percent of the students.

3. The teachers and administrators in this school are doing a terrible job because only 45 percent of the students mastered this objective.

4. The teachers and administrators in this school are doing a fine job to have gotten these students this far on this objective.

5. These students should be doing better since they come from very affluent families; the objective is simple; and the school staff is excellent.

6. These students cannot be expected to do any more than this as they are ethnic minorities, poor, suffering from the pangs of integration (or segregation or whatever), etc.

Obviously, this partial shopping list of interpretations demonstrates that the same information can be interpreted in several different ways depending on the intent of the program (i.e., are they *minimal* objectives?) and whether the tests are seen as producing a status report on the school or an evaluation (i.e., long term growth) of the school's academic progress.

There is no easy solution to this problem other than to work very hard to establish an adequate dissemination program. In this regard, the National School Public Relations Association published a useful monograph a few years ago entitled "Releasing Test Scores: Educational Assessment Programs, How To Tell The Public" which can be obtained from the association.

The Florida program provides several different school and district summaries that meet the needs of citizens, parents, educators, and legislators for different types of data. For example, the educators are interested in knowing what percentage of students mastered each of the state minimum student performance standards for each school in a district and for the district as a whole. This information is provided in computer printouts given to each district several weeks after the testing is completed. Legislators need to know the percentage of students in each district who fall below the 25th percentile (student norms) in order to provide compensatory aid monies in accordance with the severity of educational needs. Citizens need to know how their schools are doing compared to the expected minimums (in an aggregated sense). For this purpose, the department provided a "composite score" based on the percentage of students mastering the several standards. Schools falling below a particular cut-off are flagged to alert local officials that there may be a need for a closer look at the school and the difficulties it may be experiencing.

In Florida, then, there are explicitly stated standards for students which, in turn, are aggregated upward into minimum expectations for schools. There are no further aggregations to the district level in terms of composite scores. However, the implied standard is that all students will meet all minimums. This is a tall expectation, and districts would be very successful at this point if they could get 80 percent of the students to meet the minimums.

A few recommendations, then, are:

1. Establish clear purposes for the competency program and attempt to differentiate, if possible, between the assessment and evaluation purposes.

2. Build reporting systems that address the purposes of the program.

3. Differentiate the reporting systems so that they serve the needs of different audiences in the most efficient manner. Do not expect one report to serve all needs.

4. Determine whether or not there will be minimum expectations for schools, for districts, and for the state. If so, build a reporting system around it that will be understood by citizens and legislators, in particular.

5. Do not be surprised by complaints of misinterpretation of the test results. People tend to use data to meet their own needs and purposes.

9. The Incompetent: Implications for Schools

All things considered, creating and administering an assessment or competency test is easy to do. Testing technology, despite certain frailties is a fairly well established science. Technicians know how to create sets of performance objectives, match them with test questions, administer them consistently to students, process the tests with great electronic accuracy, and print the resultant data in a myriad of forms and formats. The problems begin when these results are communicated to various audiences, and the problems get serious when someone attempts to assign meaning to the data.

The action in education is not really the collection of data about achievement but in the teaching of students after the data are available. Knowing Johnny has not mastered two digit addition is important, assuming you think he should have that skill, but you still have to determine how to help him once the need has been recognized. In other words, what does one do with the incompetent?

This problem has not been completely solved in Florida although some initial steps have been taken. For example, actions of the State Board of Education have made it clear that the local districts and schools have the responsibility to assist students who have experienced difficulty with portions of the state tests. Specifically, rules of the State Board of Education make it clear that if a student does not attain one or more of the state minimum standards, he or she is to be provided additional assistance and an opportunity to be re-evaluated on that minimum standard. The student is not expected to retake the state test in the basic skills (although he or she can retake the literacy test) but, instead, must be re-evaluated through teacher observation, teacher-made tests, district-made tests, or any other appropriate method.

Grade-to-grade promotion in Florida is not tied solely to the statewide competency tests. A student is expected to meet all local and state requirements for promotion and graduation before promo-

tion to the next grade, not just pass a state test. In theory, schools would receive state assessment data on a third grade student by late November and would have the remainder of the year and perhaps even the summer term to determine whether or not the student should be promoted. The same time block is available for appropriate remediation although the student who has not attained one standard probably will not need as much time as the student who has not attained ten standards.

The state of Florida also has enacted a compensatory education program funded at a level of $26.5 million. These monies are allocated on the basis of relative need and each district receives a share. The funds must be used to supplement students' instruction and must not supplant local or federal funds already being utilized for the student. The monies go directly to students in the form of instructional services and may not be diverted for general raises in pay to teachers or other secondary purposes.

Typically, these monies have gone to districts to pay for remedial help in the form of evaluation assistance. In 1977, districts were inclined to spend their funds on the high school students who had trouble with the first literacy test. In 1978, they seem more inclined to spread the funds across the various grade levels.

It is not yet known how effective these remedial instruction programs will be. There is great hope for them but the final proof will be in terms of how many students fail the literacy test as seniors and how much progress is seen in the assessment scores over time. Clearly, there is a potential problem in that one cannot hold students indefinitely in a grade until all of the state and local expectations are reached. Some students can be expected to progress very slowly, and some may not be capable of ever reaching a particular standard. The question thus turns full circle back to the original problem of whether or not it is possible to identify appropriate minimal expectations. If educators miss the mark in identifying expectations, some students may stay a very long time in a particular grade. Or, educators may be forced into a position of watering down their post-remedial instruction evaluations to a ridiculously low challenge in order to move the studens along. Hopefully, neither of these conditions will prevail and educators will be able to find ways to help students meet the minimal expectations and move along from grade to grade at a reasonable rate. With these considerations in mind, the following recommendations are offered:

1. Do not assume that the task is finished when the competency tests have been designed and implemented.

2. Plan on appropriate remedial and follow-up activities. Determine how individual students will be assisted when they do not meet the minimum expectations.

3. Consider how instructional resources will be allocated to sup-

port remediation activities. Settle the basic philosophical question of whether or not you will "reward incompetence" through the provision of additional funds to low scoring districts and/or schools.

4. Establish a plan for tracking longitudinal progress. Be able to answer the question, "Are the students doing better today than they did in 1978 (or whatever)?"

5. Anticipate various research questions that will arise from the competency test data. Strive to build linkages with the research community in the colleges and universities that can assist in answering these questions.

6. Decide on the role and function of the agency instructional support personnel. Will they perform an audit function with low-scoring schools and districts or will they perform technical assistance roles?

7. Provide a method for sharing resources across districts and schools within a district. Everyone will be interested in remedial materials and in properly designed post-tests, and some method must be found to lessen the chance of "re-inventing the wheel."

8. Consider the implications of new remedial instruction programs for the teachers, administrators, and other staff. Are they trained to provide such services?

9. Consider the implications of remedial instruction for the basic organizational patterns of the schools themselves. Determine how such programs can be woven into current schedules and activities.

SUMMARY AND CONCLUSIONS

The purpose of this paper was to summarize how the state of Florida approaches its new competency testing programs for the public schools and to provide other educators and policy makers with suggestions for implementation of such programs based upon what has been learned thus far in Florida.

Specific recommendations were offered on such subjects as 1) defining competencies, 2) measuring competencies, 3) determining competent schools, and 4) completing remediation activities.

In the area of defining competencies, it was suggested that this should not present an overly difficult problem. Methodologies have existed for some time that are quite useful in gaining agreement on a set of minimal expectations for students. Great effort should be made to document that agreement has indeed been reached on the "minimums" as this will be a central attack point in any legal challenge. Adequate time should be allotted for this step and provisions should be made for the revision of the skills over time. The minimum competencies should be limited in scope if at all possible.

Measuring minimum competencies will always be a compromise as testing is not a perfect science, but the state-of-the-art is not so

limited that one cannot place substantial faith in the process. Tests of most basic competencies can be developed and administered with great confidence. These tests must be designed with careful documentation of the developmental process, and particular attention to protecting the tests from attack on the basis of invalidity, cultural bias, and related points. The tests should be administered early enough to provide remedial instruction but not so early as to permit students to forget the skills before graduation.

Competencies for students necessarily lead to conclusions about the competency of schools and districts. Test data are frequently used both in an assessment mode and an evaluation mode. This creates a dissemination headache for educators at all levels. By building an adequate reporting system and by clearly stating the purposes of the competency testing program, educators can alleviate some of these problems.

Competency progams inevitably identify incompetent students. Educators must determine how to assist these students and the schools in which they are found. This task has serious implications for the design of instructional systems, teacher training, school organization, and educational funding. The task of providing assistance to the incompetent is more complex than that of identifying them in the first place.

All things considered, the Florida experience thus far has been gratifying. There have been controversies, to be sure, and there will be more before the program becomes an established part of education in the state. The positive aspects of the competency program thus far are that, 1) there has been great interest in and discussion about the role and function of today's schools, 2) there is a new interest in the proper use of tests to improve student instructional programs, 3) new vigor and monies have been pumped into the state's schools, 4) citizens have a new opportunity to become involved in the educational enterprise through their participation in the development of competencies and access to public information about the achievement levels of students, and 5) thousands of students who would otherwise have been lost in the system, never to receive the assistance they need, have been identified and given another chance to learn their basic skills before graduation. The last point makes it worth the whole effort.

THE TECHNOLOGY OF COMPETENCY MEASUREMENT

William A. Mehrens

General Introduction

As Ingle and Gephart pointed out in the announcement of this symposium, competency assessment is a problematic area for educators. But we are in it and the objective for NSPER: 78 is to develop answers to ten questions on the issue. It would appear from the title given my presentation that "my" answers to the questions should be based, at least in part, on the *technology* of competency measurement. I approach this task with some trepidation but that is my problem, not yours!

Before I proceed to "answer" the ten questions let me speak very generally and briefly about the extent of competency testing, why we are in it so extensively, and what two situations I refer to when I speak of the phenomenon called "Minimal Competency Testing."

As many others have pointed out before me (e.g., Ebel, 1978), minimum competency testing has been around for a long time. A very early minimal competency exam was given when the Gilead guards challenged the fugitives from Ephraim who tried to cross the Jordan river.

"Are you a member of the tribe of Ephraim?" they asked. If the man replied that he was not, then they demanded, "Say Shibboleth." But if he couldn't pronounce the H and said Sibboleth instead of Shibboleth he was dragged away and killed. So forty-two thousand people of Ephraim died there. (Judges 12:5-6, *The Living Bible.*)

Nothing is reported concerning the debates that may have gone on regarding what competencies should be tested, how to measure them, when to measure them, how high the minimum should be, or indeed what should be done with the incompetent. We do not know the ratio of false positives to false negatives or the relative costs of the two types of errors. We do know that a minimal competency exam was given and that forty-two thousand people failed — with no chance of a retake. And some people in Florida think they have it bad!

But there have been other, less drastic competency exams, for example, the basis for certifying or licensing professionals and the basis for obtaining a driver's license.

If not a new concept, why so much fuss? Because never before has there been so much action taken at the *state* level to mandate the set-

ting of minimum competency standards for elementary and secondary students. Based on an August 3, 1978 summary prepared by Chris Pipho (Pipho, 1978a) there are 35 states that have taken some such action. It has been reported (Pipho, 1978b) that all the remaining states either have legislation pending or legislative or state boad studies under way.

Why the big push to mandate such programs? Basically because many individuals believe that available evidence suggests that (1) the quality of our children's education is deteriorating and that (2) minimum competency testing will improve the educational quality (or reverse the deteriorating process if point number 1 is true). Both points are debatable. I believe the first, some of you may not. I could quote you evidence on my side and, if sufficiently informed, you could quote contrary evidence. But this presentation is not on that topic and I will only allude to it briefly in the remainder of this paper. The second point is one about which I would prefer to reserve judgment. It will be addressed somewhat in this paper but I am not here to take a basically pro or con position with respect to the issue. I am to answer some what, how, and when questions; not whether questions.

Before getting into the heart of the talk, let me briefly state the two main variations of what I mean by minimum competency measurement. Both involve a systematic test over some basic competencies. In one case the outcome is used to control grade-to-grade promotions; in the other it is used to determine high school graduation (or type of diploma).

1. Which Competencies?

The answer to the question of which competencies should be measured in a minimal competency program is related directly to the purposes of the test, i.e., the inferences we wish to make about a person who "passes," and much less directly to the "purposes of the school." Many people apparently do not make enough of this distinction.

Although there exists a reasonable consensus about whether many characteristics are desirable or not in adults there is considerable diversity of opinion about their relative importance and the role of the school in promoting appropriate characteristics. Some people maintain that "good citizenship" or "healthy self concept" are more important in life than reading skills. Others assert just the opposite. And some who believe the former do not believe it is the primary purpose of the school to promote those characteristics. I suspect that we will never reach agreement about which charactistics we "need" in our society and the role of the school in teaching, providing, establishing, or nurturing those characteristics. That does not dismay

me, nor do I believe it should deter us from deciding on the general content for a minimal competency test. No test can be designed to assess the degree to which all the purposes of education have been achieved or even to assess whether students have achieved a level of minimal competency in all areas.

Surely no one would infer that all the purposes of education have been achieved if one passes a minimum competency test. What will people infer and/or what do we want people to be able to reasonably infer from a passing score on a minimum competency test?

Would any reasonable citizens infer, or would we want them to infer, that a passing score means the person has "survival skills" for life? Life is very varied and what it takes to survive is varied. I cannot believe the populace is so unrealistic or naive as to think in such grandiose terms. Schools do not and can not teach all survival skills. They can not even be adequately enumerated (or defined). Thus they can not be adequately measured. We should not want any survival skills inference to be drawn from a test and we should therefore not build a test to measure such non-defined competencies.

If not "survival," how about "Life Role Competencies? Porter (1977), the State of Michigan Superintendent of Public Instruction, in the foreword to a document entitled *The Michigan Life Role Competencies*, suggests that the document lists some "real-life skills that we all presumably need to have to function well in the areas of personal and family management, the civic and social areas,as aesthetic and humanistic individuals, and as employees." (I'm not sure why employers got excluded here.) The list of competencies includes such statements as:

1. Be able to project the probability of obtaining a job from among career opportunities.
2. Recognize and respond to personal physical, social, mental and emotional needs and to those of family members.
3. Perceive and use the natural environment as a basis for creative expression.

Now please understand I am *not* suggesting those statements do not represent worthy life role competencies nor am I suggesting the school has no role to play in student acquisition of these competencies. But anyone who has looked at such a list (I have only stated 3 out of 44 in the Michigan document) could surely not be very sanguine about educators' abilities to build minimal competency measures to assess each of those competencies. Should we therefore wish to encourage the public to infer that such competencies are (or are not) achieved based on some so called minimal competency exam? I think not! It would be deception if not fraud to encourage such an inference.

If we do not wish to make an inference to all life role competencies how about a subset? Which subset? How would we measure

it? And if testing programs are developed on the assumption that they assess such life-role competencies and we inform the public of such an assumption would we not have an obligation to establish predictive validity for our instruments? Could we do so? Would we be liable without such predictive validity? Obviously, I think it best if we steer away from attempting to assess such competencies. We should not try to measure, nor wish people to infer that we have assessed some adequacy in social functioning. It seems far more preferable to measure certain school or basic skill competencies. If we do not make "survival" or "effective functioning" claims we will be on safer ground.

Actually, the distinction between school skills and life skills is not a true dichotomy anyway. Porter (1977), in his foreword mentioned earlier stated that "in order to demonstrate these life role skills, students will need to have achieved certain proficiency levels within their academic and basic-skill courses."

As Brickell (1978) points out, "there are basic skills — such as reading, writing, and arithmetic — used in both school and life, which is why we call them basic." (1978, p. 589)

Brickell's distinction between a basic skill *applied* in a school subject and one *applied* in a life area is somewhat troubling to me. Suppose I ask a student to compute interest rates, balance a checkbook, fill out a job application, write a business letter, or read a bus or T.V. schedule? Why should I call these basic skills applied in a life situation as *opposed* to applied in a school subject? Certainly such tasks (subjects) are taught by teachers in schools. It seems to me they are applied *both* in school subjects and life situations. To infer schools do not teach or test for such applications does them an injustice.

Of course there probably exists a set of basic skill questions that can be identified only with *school* applications. These are probably not what we want in a minimal competency exam.

Actually I am quite happy with the Florida SDE definition of functional literacy as, "application of the basic skills to problems encountered in everyday life (Fisher, 1978). Certainly this is the general approach adopted by most states. A report of the National Academy of Education to the Assistant Secretary for Education states in part:

> Two realities underscore the importance of re-emphasizing "basic skills," assuming this phrase includes a learning core in reading, writing, and numbers. First, is the deplorable and declining quality of written composition among all sectors of young people in this country. Second, is the heavy penalty that functional illiteracy exacts . . . the Academy Panel would only underscore the proposition that the distinctive role of the school lies in the development of cognitive functioning . . . . Without the acquisition of the fundamental learning

> tools of reading, writing, and numbers, no education of value can proceed. (Report of the National Academy of Education, 1978)

But if we measure only basic skills (applied to life settings) won't other areas of school suffer? I do not think so. Remember, there is a distinction between the purposes of school and the purposes of a minimal competency test. The purpose of the latter can *never* be to assess all the objectives of school. We all know that. Of course not all skills are basic and we do not want minimums to become maximums. Few would be happy to see high school graduates who lacked maturity, self discipline, and understanding of their own value systems. But if we keep in mind the limitations of the inferences to be drawn from passing (or failing) a minimum competency test, there should not be a deleterious effect due to such limited testing.

We should not assume minimal competency standards do much at all in helping define the goals and objectives of education. They only set a lower limit of acceptable standards in certain basic skill areas. This certainly suggests that passing the minimal competency test should not be the only requirement for high school graduation. Other graduation requirements could assure breadth in other areas.

Given that my answer to the question of "which competencies" is to assess the basic skills, can I be more specific? We, or course, want any test to be content valid. Once a domain is specified, there are reasonably standard ways of building a test such that the sample of items in a test represent the domain. In specifying the domain of basic skills, one needs to keep in mind the relationship between the tested domain and what is taught in school. We should *not* be testing content not taught. On the other hand, we should not be attempting to randomly sample *all* that is taught. The tested domain must be a *subset* of materials taught in the curriculum. The domain must be defined precisely enough to rule out relatively unimportant specific bits of factual knowledge as well as processes so abstract that they appear to measure general intelligence. There should be evidence the material tested is not only taught (i.e., presented) but that almost all students are capable of mastering the material.

To get more specific in answering the question of "what competencies" would assume that the answer to question 8 (who selects the competencies) is "Me." I do not believe that is the correct answer to question 8 so I will stop here on question 1.

2. How to Measure?

There are a variety of possible meanings to the question "How to Measure." How to sample, how to administer the measures, or how to build the measuring instruments are all possible meanings. Most

people who have spoken to this question have addressed the latter point — usually specifically with respect to the *type* of measuring instrument. Let me address the other two very briefly.

The answer to the question of sampling is related to the answer to upcoming question 6 (Minimums for students or schools). Since my answer will be students, I rule out the possibility of any sampling of students. Other individuals, with the opposite answer to question 6 should reasonably enough argue for a matrix sampling of students and questions. Obviously there should be a sampling of content from the defined domains. This should be a stratified random sample using commonly accepted and widely explained measurement procedures. I will deal with this further at the end of this section.

The question of *how* to *administer* the measures could also be debated. I believe they should be administered in school, by professional educators employed by that school, taking the usual precautions for test security, cheating, student motivation, etc. In-school workshops should probably be held to help assure teacher competency in these matters.

The question of how to build the minimal competency measuring instrument is usually discussed with respect to the *type* of instrument (i.e., multiple choice, essay, performance test, product judgment, etc.). Obviously, the choice depends somewhat on which competencies one wishes to measure. Remember, I have voted for basic skills. I believe that for such competencies we can get along reasonably well with what we have traditionally called objective paper-and-pencil tests, but the answer must depend on the specific domain definitions of the competencies. Further, in general I believe the test builders should have the freedom to choose the appropriate mode of testing. (Note: I have just answered question number 10.) However, in case you reject my answer either because you have a different set of competencies in mind or because you do not believe in the paper-pencil approach let me discuss briefly the other options.

Brickell (1978) enumerates four alternatives: 1) Actual performance situations, 2) simulated performance situations, 3) school products and performances, and 4) paper and pencil tests. By actual performance he means in later schooling or on the job. This type of measurement is simply not feasible. The distinction between the remaining three options becomes somewhat blurred, especially when talking about the basic skills. Performance measures do not always get defined the same way. Mehrens and Lehmann (1975) state that a test is called a performance test if the tasks require a *manipulation of objects* rather than an oral or written response. The definition offered by the Clearinghouse for Applied Performance Testing is "a test that measures performance on tasks requiring the application of learning in an actual or simulated setting." (CAPT Newsletter, 1978) Now, if by performance they mean something physical (other than

marking the right answer), then their definition is much like the one by Mehrens and Lehmann and I reject the need for such a test format to measure basic skills. If the emphasis is on the *application of learning*, then I suspect many tests they would call performance tests I would be content to call paper and pencil tests.

There are advantages and disadvantages to both performance tests and paper and pencil tests. These are well discussed in the literature. Costs and objectivity are on the side of paper and pencil tests. Relevance (or at least perceived relevance) is *sometimes* (but not always) on the side of the performance tests.

More interesting to me than the mode of testing is how to insure that there is a match between the test items and the defined domain. I suggested earlier that the content should be randomly sampled. But how can one insure that the items *actually* measure the objective (or content) intended? One can't. But each question should be reviewed by experts in the subject matter field prior to field testing and feedback should be gotten from both teachers and students after field testing with respect to whether the items actually measured the objective intended.

Let me give an example to illustrate the problem. Suppose our objective is to determine whether first-grade children know the rank order of a set of numerals. We might ask a question such as: Which of the following numbers represents the least amount: 13, 17, 19, 21? Now the item appears to be measuring the stated objective. But it may, instead, be measuring whether first graders know the meaning of the word least! Field testing and feedback should help determine what objective the item is really measuring.

3. When to Measure

The answer to the question "When to Measure" depends (like the answer to every other question) on the purpose (or purposes) of testing. Of course the primary reason for minimal competency testing is to identify students who have not achieved the minimum. But, identify for what purpose? To help the students identified through remediation programs? To motivate students through "fear of failure"? To make a high school diploma more meaningful? Let me assume the answer is yes to all questions.

First let me suggest that there should be periodic but *not* every year testing. I believe minimum competency programs will be most cost effective if tests are given approximately three times during the K-12 portion of a student's schooling. I would suggest something like grades 4, 7, and 10. Teachers, or course, gather almost continuous data. They often have already identified those who are achieving inadequately. The formal tests will supplement the teachers' measures and confirm or disconfirm previous judgments. I believe formal

identification to be useful. Tests are credible instruments and will help motivate students (and teachers) and will help assign a minimal competency meaning to a diploma or certificate. However, I would like to stress that while I favor tests it is NOT because I believe the teachers' judgments without the tests would be grossly inaccurate.

I am opposed to every grade testing for *minimal competencies* because it is not cost effective.* Hopefully only a very few students will be identified as not achieving at a minimal level and, at any rate, those identified at grade 4 would very likely overlap considerably with those at grade 3 or 5. Further, every grade testing may result in grade by grade promotion decisions based on the results and in spite of the generally favorable press for the approach in Greensville County, Virginia, I remain somewhat skeptical about such a plan.

I selected early grades because I believe the evidence shows remediation is more effective if begun early. I believe we need high school testing for two reasons: 1) not everyone will have "made it" by grade 7 and 2) even if they had, I think it would be wise to recheck their skills. Forgetting does occur between grades seven and ten, especially if the material has not been in the curriculum in the intervening period. A good example is the ability of students to compute percentage problems such as X is Y% of Z where any two of the unknowns are given. Students in grade 7 may well be better able to do this than students in grade 10 because the concept is frequently taught in 6th or 7th grade. (Please do NOT assume from the wording above that I would ask the question in some abstract school skill mode. I would not. I would write the question as a practical life skill problem.)

Finally, let me stress that if minimal competency tests are used for high school certification or graduation there must be opportunities for students who have not passed to retake the exams. Further, no test should be used for such a purpose the first year it is given. To be fair to students there should be a phase in period.

4. One Minimum or Many?

The answer, again, depends on purpose. For example, do we wish to categorize or diagnose?

Actually, there are many related questions that could be discussed. Should there be different minimum standards for different states? Probably yes, but it depends on the answer to question 9, who sets the minimum? Should there be different standards for different

*Note this is for minimal competencies. I am *not* opposed to every grade testing with a more general achievement measure.

localities within a state? I would say no, but again it depends on the answer to question 9. Should there be different minimum standards for different students? Of course not! I think that is a ridiculous question. If the decision is to determine whether one is minimally competent in the basic skills, the same standard should define that level of competence across all students.

Now that does not mean I am opposed to reporting out levels of performance that indeed will vary across students. For example, I believe Page's (1978) approach of scaled certification has considerable merit. I will discuss this more in the next section. But I would not (nor, I'm sure, would Page) require different cut scores across different students for the same decision.

Another question pertains not to whether there should be different minimums for different students but whether there should be one total cut-off score, one for each subject matter area, or one for each objective.

Assume for the moment that two basic skill areas have been identified. If we are truly concerned with minimum performance in each area, then one can *not* use a single total score or any type of compensatory sca ing model across areas. Multiple cut-off scores are needed, at least one for each basic skill area.

But what about the sub-skills (objectives) within an area? Will a compensatory model work there, or do we need multiple cut offs? If the latter, would we require a "pass" on every sub-skill objective or only on a certain percent of them? Of course the answer, and the importance of the answer, will depend upon the covariance structure of the item scores within and across objectives and objective scores within the total test. If the test covers heterogeneous competencies, each important, then a single cut off score would not be too meaningful. However, to set separate standards for each objective would result in another problem, that of the reliability of the scores. The empirical evidence I am aware of would suggest that the covariances across objectives need to be sufficiently high so that one can defend the use of a total score within the broader basic skill. Of course if it seemed useful, one could report out objective by objective and still have only one total cut off score for the categorization decision. The usefulness of the objective by objective information would be dependent upon the reliability of each objective score as well as the reliabilities of the difference scores.

However, even if one did report out objective by objective, the total score should be based on the total number of items correct, not the total number of objectives passed. Unless differential weights were used for the objectives, the former method would necessarily be the more reliable of the two. To dichotomize objective scores before combining results in the loss of information.

5. How High the Minimum?

I would rather address the question of how to *set* the minimum rather than how high it should be. But I guess they are related and I will address both.

First off I would like to admit, like others before me, that *where* the minimum is set is arbitrary. Different methods of setting the minimum lead to different cut off scores and one can not say in the abstract that one method (or one cut off score) is superior to another. Gene Glass makes the point as follows:

> I have read the writings of those who claim the ability to make the determination of mastery or competence in statistical or psychological ways. They can't. At least, they cannot determine "criterion levels" or standards other than arbitrarily . . . the language of performance standards is pseudoquantification, a meaningless application of numbers to a question not prepared for quantitative analysis. (Glass, 1978)

So, admittedly, setting the standard is arbitrary. Further, it is politically and economically influenced. If the standards are too high and too many students fail, then there will surely be a public outcry about the quality of the schools and the unreasonableness of the standards. Further, if one is committed to remediation, the costs of remediation could be very high. If the standards are set too low then the program becomes meaningless and, if the public becomes aware of the ridiculously low standards, they will again present an outcry about the quality of the schools. The standard setters will be damned either way.

Glass raises the question of *whether* a criterion-referenced testing procedure entailing mastery levels is appropriate. He answers in the negative stating that "nothing may be safer than an arbitrary something." (Glass, 1977)

Now I certainly admire Gene Glass as a person and I agree with much of what he has said in the two articles I have referenced in this section. Further, I have been lukewarm to the criterion referenced movement, particularly its anti-norm referenced component. And indeed we might be "safer" with nothing rather than an arbitrary something. But let me for the moment take the other side. (Keep in mind, however, that I do not view myself as either arguing "pro" or "con", just how.)

There is no question but that we make categorical decisions in life. If some students graduate from high school and others do not, a categorical decision has been made whether or not one uses a minimal competency exam. Even if everyone graduated there would

still be a categorical decision if there were the possibility (philosophical or practical) of not graduating. If one can conceptualize performance so poor that the performer should not graduate, then theoretically a cut off score exists. What the proponents of minimal competency exams seem to be saying is that they believe, at least philosophically, that a level of incompetence too low to tolerate does exist; and that one ought to define that level so as to make decisions less abstract, less subjective, and hopefully a little less arbitrary than they are currently.

Now, the above is not an argument for using only minimal competency tests for graduation requirements. Nor is it an argument for using a dichotomous as opposed to continuous test score as one of the factors in that decision. What I am trying to make very clear is that ultimately, after combining data in some fashion, a dichotomous categorization exists: those who receive a diploma and those who do not. No matter what type equation is used, be it linear or non linear; no matter what variables go into the equation; no matter what coefficients precede the variable values; the final decision is dichotomous and arbitrary. The argument against minimal competency exams can *not* be that they lead to an arbitrary decision unless one truly believes that everyone, no matter what their level of performance, belongs in the same category.

Given that an observed minimal test score is going to be set, how should it be done? Theoretically this is not a problem. Decision theory tells us exactly how to proceed. One determines the "true" mastery level cutting score and the cost of false positives and false negatives. Then some simple mathematics will tell us where to set the observed cutting score (or, more precisely, how to allocate individuals to mastery states) such that the total cost of errors will be minimized. Of course we do not know what *values* to give to "true" mastery level or the cost of the false positives and false negatives!

Practically, many different ways have been suggested to set cutting scores and allocate individuals to mastery states. I am *not* going to review these here. They are thoroughly discussed in readily available literature (e.g., Millman, 1973; Glass, 1977). I do wish to emphasize that the evidence shows quite conclusively that different methods lead to different results. Further, I wish to present the following lists of do nots and dos when setting cutting scores.

Do Nots

1. Do not set cutting scores prior to constructing the items.
2. Do not set cutting scores without gathering some empirical evidence regarding item difficulty on an appropriate sample of students.

3. Do not set cutting scores without some empirical evidence regarding the teachability of the material and the educational costs associated with the instruction.

4. Do not set cutting scores without explicit consideration by representatives of parents and educators of the relative costs of false positives and false negatives.

5. Do not set cutting scores that take effect the first year the test is administered.

6. Do not conclude that more remediation is needed in one basic skill than another based on the different proportions who obtain the same percentage "pass" scores on two non equated tests.

7. Do not suggest to the public that we should be *satisfied* with evidence of *minimum* performance. [Porter, (1978), for example, in Michigan published a news release regarding the proportion of students who received "acceptable" scores. I believe this is misleading. The news release should have read "minimally acceptable scores."]

8. Do not assume that one cannot or should not *report* scores in a more continuous fashion even if some arbitrary cut off point is established.

Dos

1. Do consider using more than test information for making important decisions. If test scores are combined with other data (in a multiple regression sense), consider using the obtained raw score (or continuous scaled score transformation) rather than the artifically dichotomized value.

2. Do remember that cutting scores can and probably should change over time.

6. Minimums for Students or Schools?

It seems to me that the basic rationale for minimum competency exams is to determine whether the students have acquired the basic skills. This should result in benefit to students through remediation and also provide some minimal meaning to a high school diploma. To do this, one needs to set minimums for individual students.

Of course one could aggregate over students and report for the school but I would not see the value of setting a minimum for the schools. If aggregating over students, one must choose between some measure of central tendency which allows for compensatory effects of good students (e.g., the mean) or simply reporting the percent of students above the minimum score. I would prefer the former if the competency exam is constructed with enough ceiling such that a continuous scaled score has meaning. But that preference is based on

values, not measurement theory. It depends on what one wishes to infer from the aggregate score.

If any score is reported for the school, it should be made very clear that the attainment of basic skills by the students is not just the responsibility of the school. Parents, students, and the community at large share in this responsibility. Any reporting of a school score would be counterproductive if it resulted in reduced feelings of accountability by parents and students.

7. What To Do With the Incompetent

If we are going to spend money to identify the incompetent through testing, we surely ought to have a follow up plan. The testing *alone* will not educate the children.

There are a variety of options, the desirability of most somewhat debatable.

Schools might:

1. give students another chance to take the exam
2. encourage incompetents to drop out of school
3. not allow them to graduate
4. not allow them to receive a *regular* high school diploma
5. give everyone a regular diploma but give those who pass the exam a special certificate of attainment
6. not allow incompetent students to be promoted from one grade level to the other
7. assign such students to a less demanding curriculum or track
8. provide special instructional assistance in areas of specific weaknesses
9. make students who fail attend summer school; or
10. work with parents to teach them how to help their children learn basic skills.

Now these ten are surely neither mutually exclusive nor exhaustive options. They do reasonably represent those advocated by other writers in the field. If some of them dismay you, take heart. No one has advocated retroactive birth control! I happen to favor 1, 4 or 5, 8, 9, and 10. I am generally opposed to 2, 3 and 6 although they do have some merit.

The most recent Gallup Poll on the Public's Attitudes Toward the Public School (1978) suggests that by a majority of more than 2-1 the public would favor promoting children from grade to grade *only* if they can pass examinations. However, on a follow up question they favor (by even a larger majority) offering special remedial classes in subjects failed *rather than* having students repeat the whole year's work.

You will recall that I favor testing in about three different grades with opportunities for retakes in between. It does not seem wise to

have students who fail *repeat a whole grade*. They should receive special attention. What kind of special attention? That is a question to be answered by an instructional expert, not a measurement expert. But let me make several points:

1. A test designed to be effective in certifying competence is not an effective *diagnostic* test. Separate measures would be needed to pinpoint specific weaknesses.

2. Remediation takes time, money, trained staff, and a planned curriculum.

3. Schools should not allocate such a disproportionate amount of time, money, and staff to incompetents that the education of the vast majority of competent students is neglected.

4. Some students may never pass a minimal competency exam. For others, the costs may simply be higher than the benefits. Schools need to consider how to minimize the adverse effects of failure.

5. Hopefully if remediation is started early there will be very few students who never make it.

6. The number who do not make it is partially dependent upon the effectiveness of remediation.

7. In the final analysis, the effectiveness of remediation is dependent upon the student. As Ebel has stated, "Learning cannot be imposed. It must be pursued . . . the best a school or teacher can do is provide opportunities for learning and create conditions that make learning attractive." (Ebel, 1978)

8. To have no minimal standards in the basic skills for high school certification may well be a greater disservice to the youth of our nation than to insist on minimal competence.

8. Who Selects the Competencies to be Measured?

There are two subparts to this question: at what level is the selection made, and just who is involved at that level. Of course at whatever level the decision gets made, hopefully the opinions of the public are kept in mind. In some final analysis, the decision must belong to the public. Hopefully the selection of competencies would reflect a consensus of the constituent groups that will be affected by the tests. This includes parents, teachers, and students.

The latest Gallup Poll (1978) asked if promotion and graduation are based upon tests, who should prepare the tests, the local system, the state, or the federal government? The response percentages were local (37%), state (25%), national (28%), and don't know (10%). Gallup points out that "the public is divided on this issue, but a plurality would favor having such examinations prepared by the local school systems." (Gallup, 1978)

I think a better statement might be that the majority prefer it *not* to be the local school systems. At any rate Gallup continues as follows:

> Obviously, this is a question that needs to be debated at great length. The arguments pro and con are not well known to the general public; the answers, therefore, should be regarded as indicative but by no means final (p. 37).

Califano (1977) has opposed strongly the idea of tests and standards imposed from Washington. He suggests that if standards are federally imposed, local commitment may be lacking; that no single test is right for every school; and that control of the curriculum has always rested with states and localities rather than with Washington. I believe that if the test covers only minimal competencies in the basic skills the three objections are only minimally valid. But I suspect it is illogical to argue for federal standards under the assumption that I can set the broad parameters for those standards.

If not federal determination how about the other extreme, local? I think not. If every local district can specify its own competencies, define its own standards, and build its own tests, then for all practical purposes the standards are unknowable to most of the individuals who do not live in that community. Further, given the amount of mobility in the U.S. both during one's school and adult years it simply makes very little sense to have different standards in each local district. Further it would be extremely inefficient to have that many competencies determined, standards set, and tests built. (I probably should favor such an approach. It would be a boon to those who wish to consult!)

Those of you still alert to what I am saying can probably deduce that I favor state determination of the competencies. The state departments of education must, however, set up procedures that allow for input from various constituent groups including teachers, parents, and students.

Educators who deplore my stance and favor local control might ponder on what effect such a system could have on the state referenda that call for voucher systems.

9. Who Sets the Minimums?

This question has, like the previous one, two subparts. I argue for the state setting the minimum for the same reasons as for the state selecting the competencies. Who acts as the state's representative is a more difficult question to answer. Of course all constitutents should again be involved. But I firmly believe some measurement experts need to be involved also. (Hopefully those measurement experts would agree with my list of "Do Nots" given under the section "How High the Minimum".) While setting the minimum is arbitrary, I think measurement experts have some useful suggestions.

If Glass' (1978) description of the Florida process is accurate, they certainly could have used some technical advice. Evidently the people who set the cutting score had not even seen copies of the test!

10. Who Decides How to Measure?

Once competencies are determined with sufficient specification, the task of deciding how to measure those competencies must be left to measurement experts. This is not to suggest that others might not set certain parameters such as cost, length of test, etc. But hopefully even those general parameters are set in consultation with professional test builders.

Also, I would like to stress that once a draft test is built it is very appropriate to have panels composed of both curriculum/child development experts and lay people review the items to help insure that they do measure the objectives. A panel of test experts (other than the original authors of the test) should also review the test. Hopefully some members of this panel would be knowledgeable about and attentive to the potential for cultural and gender bias in the test items.

11. Summary

First, recall that I did *not* view my task as either defending or attacking the notion of minimal competency measurement. I tried to answer ten specific what, how, when type questions. My answers were as follows:

1) What competencies? Application of basic skills to problems encountered in everyday life.

2) How to Measure? Paper and pencil mode should suffice.

3) When to Measure? Grades 4, 7, and 10 with opportunity for retakes after grade 10.

4) One Minimum or Many? One minimum per state per basic skill.

5) How High the Minimum? There is no precise technical answer. It depends on too many variables. Among other things, it depends on availability of remediation money and the political acceptablity of failure rates.

6) Minimums for Students or Schools? Students, but it is OK to aggregate over student scores and report a school score.

7) What To Do With the Incompetent? Allow retakes of the exam; provide remediation following the gathering of more diagnostic data; do not allocate such a disproprotionate amount of time, money, and staff to the few incompetents that the further education of the minimally competent student is neglected.

8) Who Selects the Competencies to be Measured? The state with input from various constituent groups including teachers, parents, and students.

9) Who Sets the Minimums? Same answer as to question 8, only use measurement experts also.

10) Who Decides How to Measure? Experts in test construction.

References

Brickell, Henry M. 1978. "The Case for Minimum Competency Testing." *Phi Delta Kappan*, 59, 9, 589-592.

Califano, Joseph A. 1977. Remarks Before The College Entrance Examination Board Annual Meeting. San Francisco, CA. Mimeograph.

CAPT Newsletter. 1978. Volume III, Issue VI.

Ebel, Robert L. 1978. The Case for Minimum Competency Testing. *Phi Delta Kappan*, 59, 8, 546-549.

Fisher, Thomas H. 1978. "Florida's Approach to Competency Testing." *Phi Delta Kappan*, 59, 9, 599-602.

Gallup, George H. 1978. "The 10th Annual Gallup Poll of the Public's Attitudes Toward the Public Schools." *Phi Delta Kappan*, 60, 1, 33-45.

Glass, Gene V. 1977. *Standards and Criteria.* Occasional Paper Number 10. Evaluation Center, College of Education, Western Michigan University, Kalamazoo, MI.

Glass, Gene V. 1978. "Minimum Competence and Incompetence in Florida." *Phi Delta Kappan*, 59, 9, 602-605.

Judges 12: 5-6. *The Living Bible.*

Mehrens, William A., and Lehmann, Irvin J. 1975. *Measurement and Evaluation in Education and Psychology* (2nd edition). New York: Holt, Rinehart and Winston.

Pipho, Chris. 1978a. *State Activity: Minimal Competency Testing.* Mimeograph. August 3.

Pipho, Chris. 1978b. "Minimum Competency Testing in 1978: A Look at State Standards." *Phi Delta Kappan*, 59, 9, 585-597.

Porter, John W. March 21, 1978. News Release.

Porter, John W. (1977). *The Michigan Life Role Competencies.* Michigan Department of Education.

Report of the National Academy of Education. 1978. *Improving Educational Achievement.* U.S.O.E.

NSPER: 78

Kansas City, Missouri

BARRY B. BEAL
Denver, Colorado Public Schools

CHARLES FOSTER
Missouri Department of Elementary and Secondary Education

J. STANLEY AHMANN
Iowa State University

COMPETENCY MEASUREMENT AT THE LOCAL LEVEL

Barry B. Beal

It is a pleasure to be with you today. I have a special place in my heart for the Midwest, maybe because I was born and raised on the plains where a tree was something very special. Fall in the Midwest and the East is a time of rare beauty. I consider it one of nature's gifts and derive a feeling of tranquility and peacefulness from it. Kansas City, on the other hand, is a different matter.

Also, I appreciate being invited to participate in the Twelfth National Symposium for Professionals in Evaluation and Research. My purpose today is to share with you some of the practical, down-to-earth experiences we in Denver have been through in the 18 years that we have had a minimum competencies testing program.

First, I will describe in some detail the Proficiency and Review testing program as it unfolded in Denver and answer the ten questions, posed by the co-directors, relative to the development of the Proficiency and Review tests.

Second, I will describe a test development project that is now in progress in the Denver Public Schools. This project involves the development of what we call checkpoint tests. These instruments are minimum competency-like tests to be used at the third and sixth grade levels. They are, in fact, an extension downward of the Proficiency and Review tests. In the past two or three years, my office has received well over 3,000 inquiries about our Proficiency and Review test. We call it P&R for short.

Minimum competency testing is one of the primary issues in education today. The willingness with which many administrators initiate their involvement in this movement reminds me of an incident which occurred immediately after a revolution in a small Balkan country. It seems that the people wanted to assure themselves that they always would maintain their liberty. They decided that to do this, three men would have to be dispatched, the general, the mayor, and the minister of education.

The general was led up to the guillotine and placed on it. The blade was released, dropped down, but stopped just a quarter of an inch above his neck. The people decided that he could not be tried twice, so they banished him. Next, the mayor was led up, placed on the guillotine, the blade released, but it, too, stopped just a quarter of an inch above his neck. He, too, was untied and banished from the land. As the minister of education was being led up, he pointed to the track in which the blade traveled and informed them, "The prob-

lem is right here. You need a little bit of oil!"

Let us hope that we will exercise more discretion when pointing out certain educational problems (case in point, SAT scores).

Minimum competency testing as it is used here refers to the testing of a student over a specifically defined set of material and requiring a stated level of achievement before certification of mastery. Such an assessment is usually used relative to a pupil's promotion, advancement, or graduation. It is a product of the back-to-basics movement. The public is telling the schools that they want more and better teaching of reading, writing, and mathematics. The *Tenth Annual Gallup Poll of Public Attitudes Toward Education* indicated that by a two to one majority, Americans feel that children should be promoted from grade to grade only if they can pass examinations. It is clear that the public may no longer continue to be completely satisfied with traditional ways of promotion or of awarding diplomas.

One way to determine if a student has demonstrated specific competencies is to give him or her a test. Standards or mastery levels are set and the student is expected to reach them. Denver is one school system in the United States that currently operates a testing program designed to help assure that high school graduates are minimially competent in four basic areas, arithmetic, spelling, grammar, and reading comprehension. If a senior does not pass all four tests, a certificate of attendance is awarded instead of a diploma.

This testing program is called Proficiency and Review; it has been operating in Denver since 1960. However, the planning for the tests began two years earlier. The program will be briefly described and answers to the ten questions relative to the development of such a program given. It must be realized, however, that the questions are being answered in retrospect. At the time of development there was not the degree of formalization of the process that now exists.

In 1958 a public opinion survey of 400 business and industrial employers in the Denver metropolitan area revealed that questions were being raised about the validity of the high school diploma. A few graduates, it was believed, lacked competence in the basic skills. The superintendent and the board of education agreed that a demonstration of competence in the basic skills should become a graduation requirement.

1. What Competencies to Measure?

Direction was provided in large part by the community. The schools responded to a perceived need expressed by a segment of the community served.

In 1959 the Proficiency and Review tests (P&R tests) were developed in cooperation with the California Test Bureau. The tests

were designed to have four parts: numerical, spelling, language, and reading proficiency.

2. How to Test?

In view of the large numbers of pupils to be tested and the foreseen need to do some retesting and to repeat the process each year, a paper and pencil instrument was considered as most appropriate and practical.

In 1960 these tests were administered to seniors, some of whom failed one or more parts. Remedial classes were established in the high schools to tutor those pupils who had failed one or more of the tests. A voluntary summer school program was established to provide remedial instruction in the basic skills areas and the tests were administered again. Only three percent of the seniors failed to obtain a diploma, receiving a certificate of attendance instead. Two more years of experience produced approximately the same results. It was then decided to administer the tests for the first time to students during the first semester of the eleventh grade. This would provide additional time for remediation.

During the school year of 1967-68, an evaluation of the P&R tests was conducted by the research department. It was found that the tests may have been selecting on factors other than the cognitive alone. Some pupils, realizing the importance of passing this test, would "freeze" and score lower than their previous academic experience would indicate they might. As a result of that evaluation, the recommendation was made that the tests be given much earlier in the pupils' tenure than the eleventh grade so that more time could be spent helping pupils to overcome their problems.

At present, the P&R tests are administered in December and April of each year in grade nine with make-up in grades ten through twelve. Pupils who fail one or more sections are placed in established instructional programs in the basic skills to assist them in overcoming deficiencies.

3. When to Test?

As a result of eight years of experience and some careful evaluation, it was determined that the appropriate time to test pupils was in the ninth grade.

The tests are used to help teachers and administrators identify students who do not meet locally established achievement levels in basic educational skills. Once identified, these students are given remedial work or placed in review programs. The tests are appropriate for administration to students in secondary schools during the early part of the ninth grade or beyond.

Denver's school system follows a proficiency and review program routine of:

1. Initial testing of all students at ninth grade level, using *all four* of the proficiency and review subtests.
2. A period of review or remedial work for those students who fail to attain the required level on one or more subtests and are therefore considered to be deficient in one or more skill areas.
3. Retesting with an alternate form of the test. The students take only the subtest or tests that they "failed" on the initial testing.
4. A repetition of the cycle for students who continue to fail.

The individual student is the focal point of the entire program. The student is helped to understand weaknesses in the basic skills and to overcome them.

4. What to do With the Incompetent?

Denver's procedure is to identify weaknesses and reteach concepts.

5. One Minimum or Many?

If a pupil never passes all four sections of the P&R, a certificate of attendance is awarded, not a diploma.

Numerical Proficiency Test

The numerical proficiency test is designed to measure ability in the four fundamental arithmetical operations and in the basic concepts. The fundamental operations are applied to whole numbers, fractions, decimals, and denominate numbers. The test items concerned with other basic concepts involve conversion from one to the other of fractions, decimal numerals and percents; expression of large numbers and monetary amounts in numerals; computation of percentages; and determination of relative sizes of numbers in lists of fractions and decimal numerals.

The 50 items in the test are divided equally among the five areas: addition, subtraction, meanings, multiplication, and division. All items are multiple choice items and have five alternatives.

Spelling Proficiency Test

The spelling proficiency test is designed to measure the ability to recognize the correct spelling of words. All words used were checked against the *New Iowa Spelling Scale* and commonly used spelling testbooks. Most of the words can be classified at eighth grade level, although a few are easier and a few more difficult. The common

types of spelling errors are represented: doubling or failing to double consonants; the omission of a "silent letter"; vowel substitution; the "ie rule"; and phonetic spelling. The words are presented in 50 multiple choice items of four words, with the fifth choice being "none," that is, no errors.

Language Proficiency Test

The language proficiency test is designed to measure the ability to recognize correctly written english. The test contains two selections that represent compositions that might have been written by students. It is the examinee's task to proofread these selections. In the 50 item test, each line of the selection represents one test item. The student is to decide whether the line is correct or contains an error in punctuation, capitalization, or grammar.

Reading Proficiency Test

The reading proficiency test is divided into two sections: reading comprehension and vocabulary. The first section is designed to measure comprehension as reflected in the ability to recognize details stated in a reading passage and in the ability to make simple inferences. The vocabulary section measures knowledge of word meaning.

The reading comprehension section consists of three short reading passages with each followed by eight statements. The examinee is to read a passage and then decide whether each statement agrees with the preceding passage, disagrees with it, or cannot be inferred from it. The ability to recognize stated details is tested by approximately two-thirds of the items; the ability to make simple inferences is tested by about one-third of them.

The vocabulary section of the reading proficiency test contains 26 test items. Each item consists of a lead word followed by a list of four words. The examinee's task is to select from the list the one word that means the same, or about the same, as the lead word.

At present the proficiency and review tests are administered in December and April of each year in grade nine. Make-up examinations are also given twice a year in grades ten and eleven. Pupils in grade twelve who have not passed all sections of the examination are given several opportunities to make up the examination.

Parents are notified twice a year, by way of report cards, of the standing of their child relative to passing each of the four subtests. Pupils who fail are provided skill analysis reports stating the areas in which they are weak. These reports serve as bases for the development of remediation programs.

If at the end of the twelfth grade a pupil still has not successfully

completed the proficiency and review tests, special classes can be attended during summer school.

Any pupil who receives a certificate of attendance (due to not completing the Proficiency and Review test requirement) may return, at any time, to the Department of Development and Evaluation and take those subtests not passed. Once the subtests are passed, a diploma is awarded, providing all other graduation requirements have been met.

6. Minimums for Students or Schools?

Pupils in the Denver Public Schools are expected to meet a minimum standard in this program. School minimums are, however, not difficult to work into the picture. If one school, for example, has 85 percent of its pupils passing the mathematics test and another has only 35 percent of its pupils passing, some judgments about school performance can be made. These judgments, of course, must be made in light of information that goes beyond simple pass-fail ratios.

As I have previously indicated, the California Test Bureau helped develop the test battery. One major goal set for the development of the Proficiency and Review series was that about 70 percent of American ninth graders be able to answer 70 percent or more of the items on any one test.

Analyses were made of the initial P&R series which had been administered to large groups of students in grades 8-12. Then four experimental forms were developed that met not only the requirements described earlier, but also content validity and item discrimination criteria. Some items met the standards set and were retained unchanged. Others were changed to meet the standards through revision.

The four experimental forms were then administered to 1,867 eighth grade students in California, Colorado, Georgia, and Washington. Items from the four forms that met the established requirements for difficulty and discrimination power were used to develop forms F, G, and H of Series II.

For each of the four tests in P&R, the score is merely the number of items answered correctly by the examinee. Users were urged to set their own passing score or standards. There were no norms established for the tests. Again, users were advised to prepare local norms for use in their own schools or districts.

7. How High the minimum?

Denver tested a sample of pupils with the instruments and made a determination about where to set the cutting score on the basis of examining the frequency distribution of scores and identifying scores that reflected a separation of the distribution into a low scoring

group and a high scoring group. Over the years this score has moved upward somewhat and is now set at 32 out of 50 items on each subtest.

8. Who Selects the Competencies?

In Denver's case the competencies to be measured were defined broadly by the results of a citizen survey. They were refined by teachers, academic supervisors and administrators serving on a committee whose responsibility was to oversee the development of the test battery.

A most original and fascinating set of competency based requirements was developed by the schools and citizens of Craig City, Alaska. The town is located on an island that serves a wide area having rugged climate and terrain. Along with usual requirements in reading, writing, and math, Craig City citizens added competencies appropriate to their needs. High school graduates must be able to float for two minutes, perform mouth-to-mouth resuscitation, read a marine chart and answer questions relating to directions, water depth, rock and reef identification, and aids to navigation. Craig City's requirements may be the best example in existence of combining local and national skills necessary for success in adulthood, and/or survival.

9. Who Sets the Minimums?

Minimums were set in a pragmatic manner by school personnel. Those involved in making recommendations were teachers, academic specialists, principals, and central administrators working as a committee. The final decision, of course, was in the hands of the superintendent and the board of education.

10. Who Determines How to Measure the Competencies?

In Denver's experience, once again the decision to use a paper and pencil approach was made by the testing committee for practical reasons.

By way of summary the answers to the ten questions relative to Denver's experience are:

1. What competencies are measured? Basic skills consisting of reading, arithmetic, spelling, and language usage ae measured.
2. How to measure the competencies? A paper and pencil instrument is used.
3. When to measure the competencies? Competency is measured at the ninth grade level.
4. One minimum or many? Only one minimum.

5. How high the minimum? Minimum passing level is 32 items correct out of 50 on each test.

6. Minimums for students or schools? Minimums are for students.

7. What to do with non-mastery pupils? Identify specific skill weaknesses and reteach concepts.

8. Who selects competencies to be measured? The community and professional staff selected the competencies.

9. Who sets the minimums? The professional staff set the minimum.

10. Who determins how to measure the competencies? The professional staff determined the measurement method.

In general, we are pleased with the results of this approach and are now in the process of developing "checkpoint" tests to be given to children at the end of the third and sixth grades.

These tests, like the P&R, will be tied to instructional objectives in the areas of reading, arithmetic, spelling, and language usage and will be integrated with the P&R testing program. They will be used to identify pupil strengths and weaknesses. Pupils who have difficulty in any of these areas will be provided with special assistance to overcome their problems.

Answers to the ten questions relative to the "checkpoint" tests follow.

1. What competencies should be measured? Reading, arithmetic, spelling, and language usage will be measured.

2. How should competencies be measured? Paper and pencil instrument will be used.

3. When should competencies be measured? Competencies will be measured in third and sixth grade.

4. One minimum or many? One minimum will be used to determine if a pupil needs additional help (remediation). A higher cutting score may be used in conjunction with emerging district programs for gifted and talented pupils.

5. How high the minimum? Not determined at this point.

6. Minimums for students or schools? There will be a minimum for pupils.

7. What to do with non-mastery pupils? Identify of specific skill weaknesses and reteach concepts.

8. Who selected competencies? Professional staff selected the competencies based on the expressed concerns of citizens regarding basic skills instruction.

9. Who sets the minimums? The professional staff committee named to oversee test development will make recommendations with advice and consent from the superintendent and board of education.

10. Who determines how to measure the competencies? The professional staff committee composed of teachers, academic

specialists, principals, central administrators, and research and testing personnel.

Testing programs such as this one can serve a variety of purposes. But mostly, they should be used to help pupils. They should be as positive in nature as possible. We are all aware of the possibility of misusing tests and test data. A school district contemplating a minimum competency testing program should strive to minimize potential misuses or abuses and to maximize the benefits to pupils involved.

A CASE STUDY OF BASIC SKILLS TESTING BY THE MISSOURI DEPARTMENT OF ELEMENTARY AND SECONDARY EDUCATION

by Charles Foster

During the meeting of the Missouri State Board of Education in October, 1975, the question was asked, "How do Missouri students perform?" I responded that Missouri students do quite well. Each time they are compared with a national norm, they score above the national average. This same question was repeated at each board meeting that I attended during the next five months.

The state board members asked me to attend their April 1976 meeting. They asked that I do some homework in the whole area of student performance on tests. I looked at the research done at the University of Texas that reported that nearly 20 percent of the adults in our nation can be described as illiterate and that only slightly over 46 percent are functioning proficiently. This report was given to the state board because the reports currently making the rounds would lead one to believe that schools are deteriorating but if one were an adult, all was well. Data were presented to the board about the results of tests of Missouri students in reading and mathematics on fourth, sixth and ninth grade tests to demonstrate that the students were performing effectively.

Missouri student performance was then compared with national data. Again it was evident that the Missouri scores were higher than those of the nation as a whole. It was also evident that the Missouri scores were declining at the same rate as national scores. The state board members reactions to these data led to editorials and other media reports about the declining performance of students on the Scholastic Aptitude Test.

	VERBAL		MATH	
	Missouri	*National*	*Missouri*	*National*
1971-72	483	453	512	484
1972-73	465	445	509	481
1973-74	475	444	510	480
1974-75	465	434	500	472
1975-76	464	431	505	472
1976-77	464	429	506	470
1977-78	465	429	506	468

Several members of the State Board of Education had received letters from citizens across the state. They came from parents, students, industrialists, farmers, businessmen, newspaper editors and teachers. These letters did not relate to the board areas of educational offerings but expressed concern about the performance of students on what is called everyday application of the basic skills. Listings of these skills have appeared frequently in news stories. The news reports said that students could not make change, could not balance checkbooks, could not read a menu, could not spell, could not write sales tickets, could not write complete sentences, and on and on.

At their evening meeting in April 1976, following the discussion of student performance, the State Board of Education established educational priorities for the state. The number one priority was "basic skills learning." The Department of Elementary and Secondary Education was subsequently directed to do a comprehensive study and to develop a test instrument to evaluate the students' performance in the application of the basic skills to everyday situations.

1. What Competencies Should be Evaluated?

The State Board of Education stated that each student should:

1. be able to read at a literate level.
2. be able to handle the English language in written and spoken form.
3. be able to apply arithmetic to everyday life situations.
4. know the fundamental processes of government essential to good citizenship.
5. understand the basic economic principles that affect everyday life.

The Commissioner of Education appointed seven people in the department to head up the study and the development of an evaluation instrument. The committee members represented special education, adult basic education, statewide testing, statewide assessment, curriculum, curriculum development and curriculum implementation. The committee started interviewing superintendents, principals, and teachers to determine if, in fact, some students who could not read, figure or write were graduating from school. All those interviewed agreed that there were students receiving diplomas who could not perform in the "basics" as called for by the State Board of Education. Estimates of the percent of such students ranged from 5 percent to 20 percent.

The committee studied materials for evaluating student performance from other states. It examined the objectives of various testing programs. Members of the committee talked with subject matter specialists, business and industry managers, college and university officials and students. The committee developed a role and function

for a state advisory council. The state advisory council included representatives of every educational group and association in the state, students, parents, news media, business, labor, agriculture, legislators and school board members. Sixty people were invited to participate on the advisory council.

On July 29 and 30, 1976, the advisory council met with the department committee. They studied the rationale and concepts developed as bases for test development and the objectives that had been prepared in draft form. The council made recommendations to the committee about further development of the test.

An interim report was then prepared for the State Board of Education. We asked, "Was it their intent to make everyone equal?" The answer was no. "Did they want equal educational opportunity for every student?" The answer was yes. "Did they want to emphasize skills a person must have in order to survive?" The answer was no. They wanted to test basic skills, which were defined as skills important in managing everyday life situations. The board said the need was to identify those students who had not really learned these basic skills and could not apply them in real life situations. Further, they agreed that schools needed to check on each student's progress in skill development at regular intervals and do prescriptive teaching in the areas of weakness.

Meaning of the High School Diploma

Our committee gave long and serious study to this question. All members of the committee were in agreement about what the diploma stood for in 1952. It stood for "academia." High school graduates in 1952 could write a good paragraph. They could attend college without the necessity of remedial courses. Of course, only 26,000 graduated from the public schools of Missouri of 88,000 students who had started the first grade twelve years earlier, in 1940.

In 1952 state legislation was passed that required students to attend school until they were 16 years of age. Recently, legislation was passed that required the schools to provide a program that would meet the individual needs of each student. The net result is that 64,000 students graduated from the public schools in 1977. Twelve years earlier, in 1965, 89,000 students had entered the first grade.

The legislation in essence had made it illegal to fail students. The schools were required to provide for them. They could not turn students who failed out into the streets as had been done in the past. The meaning of the high school diploma has changed. Now, if a student is a special education student and completes the prescribed course of study, he/she is entitled to a diploma. The transcript has become even more important than the diploma because the schools

no longer exclusively offer college preparatory programs. A student is required to attend school, even if the student is an unwilling learner. Because of this, our committee recommended to the state board that passing the basic skills test should not be a requirement for graduation.

Public Support

During October and November of 1976, more than 3,000 Missourians attended Regional Educational Conferences held at 11 sites across the state, to discuss competency-based testing and the development of the Basic Essential Skills Test (BEST). At each conference, department of education staff members gave presentations about the concepts and objectives of the BEST. Then, in small group sessions, the participants discussed the pros and cons of the concepts and objectives.

At each regional conference, the participants (who included educators, parents, students, farmers, businessmen and legislators) were asked to complete a questionnaire expressing their opinions and providing information for the State Board of Education. Of those who responded to the questionnaire, 83 percent were in favor of the concept of a basic skills test. A majority of 67 percent favored the idea of requiring all schools to give the test. On the question of requiring the test for graduation 56 percent expressed opposition, 21 percent favored the requirement and 20 percent were undecided.

Sixty-seven percent also favored giving the test for the first time near the end of the eighth grade. Only seven percent of the respondents gave suggestions for other times when the test might be administered. Those suggestions ranged from the second to the twelfth grade.

After considering the data gathered at the regional conferences, the State Board of Education took action at its November 1976 meeting. The board decided that the BEST would be available on a voluntary basis the first year, and that schools in Missouri would be required to administer the BEST to all eighth graders after July 1, 1978. Thus, it would become mandatory statewide in the Spring of 1979.

Test Development

In September 1976 subject matter specialists were identified and invited to participate with the department staff in test development. The subject matter specialists were convened for two day sessions on two occasions during September 1976. Schools were selected and their cooperation was secured to participate in field testing of the test items that had been developed. The field testing of the items occurred between November 15, 1976 and December 15, 1976. Reac-

tions of both the students and teachers were secured from each of the 51 schools that participated in the field testing.

Between December 15, 1976 and January 1, 1977 the tests were scored and the data analyzed. On January 6-7, 1977 the subject matter specialists were reconvened to assist in additional analysis and in the selection of the field test items to be used in the pilot test. On January 24, 1977 a random sample of schools was selected to participate in the pilot testing. On January 31 and February 1, 1977 two measurement specialists reviewed the test development to date and made specific suggestions for test development and about pilot testing procedures. The piloting of the basic skills test was completed in a stratified random sampling of public schools between the dates of March 15 and April 15, 1977. Following data anlaysis the Basic Essential Skills Test was printed and made available to the public schools of the state to test eighth grade students. Missouri has 76,000 eighth grade students. Sixty-one thousand of them were tested between the dates of March 13 and April 28, 1978. The test is to be mandatory for all public schools between March 1 and March 15, 1979.

2. How Will the Basic Skills be Measured?

Before one decides how the basic skills are to be measured, you must decide what kind of information is desired. The advisory council, the regional conference participants, and the State Board of Education all agreed that the testing should be done in part with criterion-referenced objective-based pencil and paper instruments, with some objectives that could not readily be tested by pencil and paper assessed at the local level. Suggested criteria for assessing the objectives at the local level were developed and provided to the schools. The schools could use the suggested criteria or develop their own criteria for the assessment.

3. When Will the Basic Skills be Measured?

In Missouri a statewide testing program is provided that makes tests available in kindergarten, grade 1, 4, 6, 9 and 12. Tests are being developed for grades 2, 3 and 5. The tests for grades 2 through 5 are all criterion-referenced in the areas of reading and mathematics. They measure academic skills and are sequential in nature. The sixth grade test is a combined norm- and criterion-referenced test in reading and mathematics. The criterion-referenced portion of the test is an everyday application of the basis skills.

It was decided that Missouri would test the basic skills at the end of the eighth grade. This is generally the last grade in which formalized reading is taught in Missouri schools. There has also been a

general tendency to teach algebra in the ninth grade. By testing in the eighth grade, students who need additional instruction in reading and basic math can be identified and courses can be developed. In addition, areas of need in citizenship and economics can be identified and instruction provided for specific students.

4. How Many Minimums Will be Set?

This question must be faced and definite decisions made. If the test is challenging to the more able students, then the less gifted will probably be so confused that they will not attempt to perform on the test and will give up. Typically, inner-city students perform less well on standardized achievement tests than do those students in wealthy suburban school districts. Should a separate standard or minimum be established for inner-city students? Would this be a form of discrimination if lower standards were established for inner-city students? Could parents be fooled into believing that all is well in terms of learning when in fact their child was not learning to read, write or compute. What about the student who graduates and then can't perform the most basic operations required in the world of work? Have the concerns of the adult world been about the number of graduates or about those people who have graduated and cannot read and write and, as a consequence, have a very real possibility of being on welfare for their entire lives?

As the questions were asked in Missouri, it was decided that the issue was to identify each of the students who could not test out at one base or standard. It was further decided that minimums were not to control schools or their offerings. The only exception to the minimum would be for those students who are in the eighth grade and who are classified as special education students who have an Individual Educational Program (IEP) required by P.L. 94-142. These students can be exempted from the Basic Essential Skills Test, or take parts of it, or have the administration altered in some way. The determination is done at the local school level and written into the IEP.

In this discussin of minimum performance levels, we should keep in mind that the BEST was not designed as an achievement test to find the strengths or weaknesses of all students in the areas tested. As a group, Missouri eighth graders "top out" on the test. For example,

10% of the students made a perfect score on the Math subtest

15% made a perfect score on the Reading subtest

8% made a perfect socre on the Government/Economics subtest.

In addition, about one-third of the students missed no more than two items on the Math subtest; more than half the students missed no

more than two items on the Reading subtest; and, about one-third missed no more than two items on the Government/Economics subtest. So you see that we haven't developed a really tough test. We are, in fact, dealing with skills that the great majority of students can apply to everyday situations.

If the BEST deals with minimum performance, has the department of elementary and secondary education therefore failed to provide for finding the performance levels of all students? A variety of achievement tests are available commercially. These tests do provide measures of student performance in a wide range of subject areas. In addition, the department has developed an assessment program that evaluates student performance in all areas of the school curriculum.

This assessment program tests performance in four educational goal areas established by the State Board of Education. They are intellectual development, social development, physical development and career development. The four goal areas are further broken down into seventeen subgoals. Approximately 150 objectives are tested. The test provides group data for districts that wish to review their instructional programs. Because the assessment does not attempt to measure minimum competencies, the group data gives a school a good measure of the level of student performance in the areas tested.

In response to all of the publics that helped formulate the overall design of the BEST, it was decided that a single set of minimums would apply to all students. This did not mean that the school had no responsibility for those students who failed. In a few moments I will discuss programs for students who fail.

5. How High Will the Minimums be Set?

In setting the levels of performance needed for passing the BEST, some other questions must be answered. At what grade level will the test be administered? Will the students have other opportunities to pass the test? Will there be any penalties for failure, such as withholding the high school diploma?

As stated previously, it was decided that the Missouri BEST would be administered in the spring of the year to all eighth grade students with the IEP exceptions. It was further decided that those students who fail would be retested in the following year.

Because the concept of the BEST also embodied the philosophy that the testing should be of help to teachers for identifying some specific instructional needs of certain students, the department developed teacher activities booklets that list all the objectives and provide several suggested instructional activities for all grades, K-12. In addition, it was believed that parents should receive reports on the performance of their children for each objective.

Administration of the BEST in the eighth grade gives students who fail an opportunity to retake the test as many as four more times before graduation. The State Board of Education has decided not to withhold the diploma from those who fail the BEST at the present time. The people in the regional educational conferences voted 56 percent against reguiring the passage of the BEST as a prerequisite for the high school diploma. The BEST committee members of the department of elementary and secondary education have recommended that passing the BEST not be a prerequisite for receipt of a high school diploma.

Based on these considerations, we reached the following decisions. The BEST would be developed so that it covered basic skills that most people agree are important to enable a person to better manage his/her everyday life. Eighth graders should be able to answer the test items by virtue of having completed eight years of schooling. The exception to this would be those students for whom an IEP is required.

The test was developed with the intent of measuring whether a student had mastered a set of objectives. The intent was not to test vocabulary. It was decided that the test would have three test items for each objective. The items in the paper and pencil portion of the test would be multiple choice with four possible answers. The minimum was set in the following way. Each student must answer at lest one item correctly for each objective and make an overall score of 75 percent on the test. Each school was directd to establish its own minimums for the ten objectives assessed at the local school level.

6. What Should be Done About the Incompetent?

The basic skills testing in Missouri this spring (1978) revealed that 10.8 percent of the students had failed the reading/language arts test, 32.6 percent had failed the mathematics test and 26 percent had failed the government/economics test. Sixty percent of the students had passed all three tests. On the locally-scored objectives the failure rate was approximately 20 percent for each objective.

The most important report on test results is for the individual student. When a student fails it may not mean that the student is incompetent. It is possible for a student to fail even though all but three of the items on a subtest have been answered correctly. On the other hand, it would be possible for a student to pass who had missed nine of the items. The student who barely passed may be in much greater need overall than the student who missed only three items.

There has been some concern expressed about the way the test is scored and failure established because of the possibilities just mentioned. The philosophy used in setting the passing criteria is the same as that for passing a driving test. Each objective has been determined

to be important. For example, if the applicant failed the parallel parking part of the driving test and passed everything else, the applicant would still fail and be required to repeat the exam at some future time.

All of the people who worked on the development of the BEST believe that it is important for each school to have an instructional management system that identifies for each student and his/her parents how the child is doing and what the problems are, if any. Parents should be involved in positive activities to help the child achieve success.

Ideally the schools should engage in diagnosis of each student who fails. Following this, instruction in the areas of need should take place. The following year, the student retakes the subtest that was failed.

The department personnel followed a plan of people involvement — professional educators from all areas as well as citizens. More than 10,000 educators and other citizens, including students, contributed to the selection of competencies to be measured. This, of course, is time consuming, but when something as important as the future of our youth is concerned, the effort is worthwhile. We believe the schools belong to the people and should be providing an educational program that guarantees our future as best we understand what we want and need.

7. Who Sets the Minimums?

The Missouri legislature passes laws that govern our schools. They have given certain rights to the State Board of Education, one of which is the right to establish standards for schools to maintain quality. This includes the right to determine minimum standards for graduation from high school.

8. Who Determines How to Measure the Competencies?

In answering this question, it must first be understood that the vast majority of the competencies are determined and measured at the local school level. The remarks here concern those minimum competencies that have been identified as needing to be measured for every student attending and graduating from the public schools of the state. Education has been a state function with local communities given local control with certain laws, rules and regulations for their governance. Because of this, it is believed that the how of measurement should be determined by as many of the people as possible. In Missouri, the greatest single source of information on student competence has been the teachers. They have the students in the classrooms. They know first hand the constraints and the limitations

of the schools. The perceptions of teachers must be tempered by reactions from the citizens who pay for the schools. After gathering information from all the publics, the State board of Education examined those instruments that had been developed and established a policy as to how student performance would be measured.

Even with all of this effort, there is still misinformation and lack of understanding. There are still those individuals who would have done it differently and some who would not have done it at all. It is a complex problem. Testing is not a panacea. In any testing program, the use of test results determines whether or not the testing has been worth the effort. The jury is still out on that issue for the Missouri Basic Essential Skills Test.

BASIC ISSUES CONCERNING COMPETENCY-BASED TESTING

J. Stanley Ahmann

Competency-based education and the testing associated with it are hardly new. Indeed, they existed in full flower well over a hundred years ago in England (Glass, 1978). Following the Crimean War, the British Revised Education Code of 1861 was adopted. According to this code, schools were directed to educate students to meet specified standards of performance in reading, writing, and mathematics. A system of external examiners was established and all final-year students were tested to ensure attainment of minimal levels of efficiency in these learning areas. This accountability program lasted about 20 years before being overturned by its critics, one of whom was the poet Matthew Arnold.

From that day until now, efforts of various kinds have been made to establish the competencies required for living in an adult world, and graduating from secondary school. In spite of these efforts, or maybe because of them, a number of crucial issues still remain unresolved with regard to minimum competency testing (Brickell, 1978). No doubt the three outstanding issues are as follows:

1. What competencies should be required?
2. How will these competencies be measured?
3. Is it possible to establish defensible minimum levels of attainment for each competency?

Only if there is agreement with regard to the answers to these questions for a specific educational setting can intelligent action be taken. In other words, solid answers to these questions are needed in the planning and development of a minimum competency testing program if the interpretation and use of results are to meet the requirements of both students and educators.

1. Competencies to be Required

An analysis aimed at determining the competencies to be required usually begins with an examination of the core elements of the adult culture. To the greatest possible degree, efforts are made to identify the essential competencies required to perform life-role activities successfully. It is reasoned that these "life competencies" reveal the "school competencies" that a student should actually acquire before he or she graduates from secondary school. Minimum levels of such competencies could be acquired by mastering discrete cognitive, affective, and psychomotor skills basic to each. When integrated and

adapted to a particular social context, they serve as the building blocks on which competencies ultimately depend (Spady, 1977).

Some theorists believe that humans have a biological predisposition to be proactive rather than reactive to their environment. Furthermore, if humans are competent they can successfully manipulate their environment, making demands upon it as it makes demands upon them. With this in mind, a concept of competence has been developed (Block, 1978). Basic to this concept is the belief that competence reflects the capacity of humans to interact effectively in three kinds of environments, each of which is designated by particular adult roles. These roles are the ones which society would normally assign, those to which one might reasonably aspire, and those which one might invent for oneself. These are known respectively as "socially-ascribed" environments, "self-selected" environments, and "self-developed" environments, as shown in Figure 1.

Figure 1
A Concept of Competence

(From Block, 1978, p. 13)

Also shown in Figure 1 are the three domains, intellectual, motor, and emotional, where mastery may be required. Using this diagram Block defines competent students as "those who can, and want to interact effectively with the three kinds of environments posed by the socially-ascribed, self-selected, and self-developed roles they face upon graduation from high school" (Block, 1978).

This concept of competence does not reveal the particular competencies that should be considered in competency-based education, but does suggest a manner of thinking about them. It is immediately clear that when trying to determine school competencies, we first need to be seriously concerned about what it means to be an adult. It

is also clear that the mastery required spans the three domains so often cited, namely, the cognitive, affective, and psychomotor domains. Furthermore, the degree of complexity of the tasks students must master varies from some of the most simple learnings to those that are extremely complex. Finally, if attaining school competencies is necessary to smooth the path to adult roles, then schools must educate the student for both today and tomorrow. In view of the rapid rate of change that exists in our society, this is extremely difficult if not impossible.

Identifying adult competencies and relating them to school competencies is a very complex task that we have failed to handle adequately. More help is needed from both sociology and psychology in determining the general demands that society places upon adults and the roles adults must play as they interact with the environments mentioned. Existing literature does give us some insights into this matter, as do state, regional, and national surveys that sample the opinion of various reference groups about adult competencies required today.

One of the more ambitious efforts to identify the specific skills students need to enter the adult world was conducted by the Adult Performance Level Project. Approximately 274 competencies were identified, representing five broad areas: occupational knowledge, government and law, health, consumer economics, and community resources. Also, four skills needed by adults to function effectively in society were established: communication, computation, problem-solving, and interpersonal relations skills.

A national survey based in part on these findings attempted to identify those competencies that various segments of the national population believed to be important for functioning adults in our society and to ascertain which competencies should be taught in the schools (Bossone, 1978). Students, parents, teachers, administrators, school board members, employers, and legislators were surveyed. Among other questions, the respondents were asked to indicate the degree of importance to the school program of 54 competencies from the Adult Performance Level Program. An overwhelming majority of the individuals surveyed indicated that minimum competencies should be taught in the schools, but that this teaching effort should not take precedence over the teaching of traditional academic skills such as reading, writing, and arithmetic. There was an important amount of disagreement as to which specific competencies should be taught.

The National Assessment of Educational Progress also has identified basic life skills by using panels representing various reference groups. These skills are:

1. Personal finance and consumer protection skills
2. Health maintenance skills

3. Interpersonal skills
4. Family responsibility skills
5. Career development skills

Various exercises have been developed and used to measure degrees of achievement of these skills by 17-year-olds (National Assessment of Educational Progress, 1977, 1978a).

In spite of the foregoing, existing minimum competency testing programs typically place heavy emphasis on the basic skills that lend themselves to paper-and-pencil testing with objective scoring. In other words, reading, mathematics, and language skills are tested with little attention paid to other learning areas, such as listening, speaking, science, and social studies. While it is clear that these areas are crucial to the development of adult competencies, it is also very clear that they represent only a beginning point in the determination of the full spectrum of adult competencies needed.

2. Measuring Competencies

Because achievement assessments have various purposes, it is necessary to look at different strategies for measuring achievement in order to determine which is the best for a given situation. Three major strategies are available (Millman, 1978). The first is item-centered, the second is objective-centered, and the third is subtest-centered. These strategies differ widely in terms of the scope of the achievement domain measured and the quality of the measurement made.

When the item-centered strategy is used, countless items are scattered throughout a broad achievement area. Very often only one item is used to measure the degree to which a student achieves a designated objective. Hence it is possible for the total test to be based on a wide variety of objectives. Since it is anticipated that the responses to each item will be interpreted separately, we obtain a comparatively limited amount of information about the achievement of a great number of objectives. This is the basic strategy used by the National Assessment of Educational Progress.

Due to the unreliability inherent in the measurement of an objective by means of only one or possibly two items, the item-centered strategy is not recommended for measuring competencies on various specific skills. On the other hand, the computation of an educational achievement index for a broad achievement area (for example, science) represented by the total collection of test items is suitable (Millman, 1978).

The objective-centered strategy is based not on hundreds of objectives, but perhaps a few dozen. Several items are used to measure the achievement of each objective, perhaps as many as 10 to 20. As a result, the reliability of the measurement has increased and yet a fair-

ly wide variety of objectives is included within the testing effort.

The subtest-centered model is based on comparatively few skills (less than ten), each of which is measured by a large number of test items. Many commercially available norm-referenced test batteries are based on the subtest-centered strategy.

Each of the foregoing strategies has certain desirable features, depending upon the purpose of the assessment. If one merely wishes to have a gross summary index for a large achievement area, then the item-centered strategy is probably best. On the other hand, if one wishes to have data for diagnosing student achievement problems, then the objective-centered strategy is no doubt better than the other two (Millman, 1978). On balance, the objective-centered strategy is probably best for minimum competency testing. Given that such programs claim to be designed to identify students who are not achieving successfully and to direct them to suitable remedial programs, then the objective-centered strategy would offer considerable information about achievement weaknesses in a reasonably reliable manner. The item-centered strategy seemingly provides more diagnostic information but does so in quite an unreliable way, whereas the subtest-centered strategy fails to provide definitive information and hence is of little use for diagnosis.

Paper-and-Pencil Testing With Objective Scoring

No matter which of the three strategies is selected for a minimum competency testing program, the tests typically are composed of paper-and-pencil items that can be scored objectively. Furthermore, most programs follow criterion-referenced test building methods rather than norm-referenced methods. In view of this, there is concern about the limitations inherent in the use of objective test items. Identifying and describing the domain used for criterion-referenced testing are most difficult tasks. Certainly the manner of stating the objectives selected is crucial to successful test construction. Specifically, the action verbs used in the objectives must define the manner in which the information is gathered to determine whether students have achieved these objectives. For instance, the action verb may require the student to describe, construct, compose, measure, or analyze. Are these the activities that are required by the typical objective test item? In contrast, many such test items are based on recognition of a correct answer within a group of possible answers that are not correct. Although the ability to perform this task may be a desirable outcome of the educational process, the fact remains that it may be a different skill than the one required by the stated objective.

Consistent with the foregoing is concern about the degree to

which objective test items measure the ability to recall information rather than the ability to use information in a novel, realistic situation. Much has been written about the former and considerably less about the latter (Ahmann and Glock, 1975). A number of exciting and clever efforts have been made to test for more than memorization and still maintain objective scoring. These are difficult, however, and do not necessarily yield easily interpreted results. On the other hand, it must be said that this type of testing is notably absent in minimum competency testing as presently practiced, even though it is eminently reasonable to expect that at least part of such tests be based on situations resembling the real world of the adult.

Performance Testing

A little used but important component of minimum competency testing is performance testing. A performance test is one that measures performance on tasks requiring the application of learning in an actual or simulated situation (Slater and Nafziger, 1978). Such testing efforts reduce the artificiality in the testing situation and heighten student interest. Driving and swimming tests are everyday illustrations of this type of testing. It is also found in the school classroom in areas such as science, physical education and health, music, art, and foreign languages.

The relative importance of performance testing in minimum competency testing is determined by the nature of the competencies involved. Consider the following, which are included in the Adult Performance Level Program:

1. Writing a letter to a community agency requesting written information regarding the services provided.
2. Reconciling checking deposits and checks written to calculate the balance in a checking account.
3. Registering to vote and following established voting procedures.

The foregoing represent adult competencies which, to be measured, will require the study of student procedures and products. Evaluation of procedures and products creates all kinds of methodological problems that are difficult to handle in large-scale assessments. Nevertheless, it is unfortunate that these efforts are made so infrequently in minimum competency testing. It is better to endure some of the problems associated with this type of assessment and live with the consequences, than to avoid it completely and thereby allow a large degree of artificiality to exist in the testing program.

3. Setting Minimum Levels

There is little agreement as to whether it is defensible to establish minimum levels of competence and to interpret these rigidly. The methods used have been repeatedly criticized and praised, more often the former than the latter. Furthermore, there is disagreement as to whether more than one level should be established for each competency, and whether these levels should be for students, for schools, or for both.

One school of thought holds that educators, as professionals, have internalized standards with regard to educational competencies, and that it is the professional responsibility of each to identify these standards and therefore establish minimum levels. It is argued that practical decisions of this nature are being made by educators all the time. For instance, teachers must decide when to stop teaching a particular subject or unit. One way of deciding whether sufficient instruction has taken place is to establish levels where "acceptable" performance is thought to have occurred.

Some express this point of view quite strongly (Clearinghouse for Applied Performance Testing, 1978). For instance, Dwyer stated, "Use of judgment in setting standards is a necessary and desirable intersection of social values and measurement technology. The setting of standards is as much a social and political process as a psychometric one. . . . In other words, judgment doesn't contaminate the standard setting process; it strengthens it." Several methods based on this idea are in use (Zieky and Livingston, 1977).

An opposite point of view has been presented with equal vigor (Glass and Smith, 1978). The central theme of this position is that human judgments are inadequate simply because humans make mistakes. In the case of minimum competency testing, these mistakes could have disastrous effects on the lives of young Americans. Glass and Smith believe that the vagaries of learning and measurement are so little understood that *a priori* specification of performance standards is foolhardy.

Six possible methods of determining a criterion score for criterion-referenced tests have been examined by Glass and Smith. They are:

1. Performance of others
2. "Counting backwards from 100%"
3. Bootstrapping on other criterion scores
4. Judging minimum competency
5. Decision-theoretic approaches
6. "Operations research" methods

All of these are found wanting in one regard or another. The method using the performance of others to establish minimum competency levels is actually a norm-referenced method. For example, the

mastery level on a test may be established as the median test score earned by a certain type of student. The second method, "counting backwards from 100," is based on the idea that the desired performance level is actually 100 percent; it is recognized that perfection is impossible, however, and hence a 5 or 10 percent loss is allowed. According to the bootstrapping method, the test in question is articulated with an external designation of "success" or "mastery." The method of establishing this external cutoff point is typically quite arbitrary and consequently the value of this method is in doubt.

Using expert judgment to determine minimum competence is repeatedly questioned. In the first place, the concept of competence is difficult to define. Does it mean the "least permissable" level of skill? If so, where does this point fall on a continuum stretching from absence of the skill to excellence with regard to the skill? In the opinion of Glass and Smith the idea of minimum competency is "bad logic and even worse psychology" (Glass and Smith, 1978).

The decision-theoretic approaches and the "operations research" methods are the most technical. Usually they rely heavily on sophisticated statistical analysis in order to determine cutoff scores for criterion-referenced tests. Both meeting the assumptions of the statistical analysis and gathering extensive data prevent these methods from being used widely, and consequently they have a somewhat modest impact on the minimum competency testing movement as it is now practiced.

Glass and Smith conclude that the only practical criterion that is safe and convincing in education is change in levels of achievement. Since we are incapable of making satisfactory absolute judgments of performance, we should make relative ones. An improvement in performance is pleasing and a decline is not. Therefore, "goodness" and "badness" should be replaced by "better" and "worse." In this connection, it should be noted that a variety of changes in levels of student achievement are occurring regionally and nationally in fundamental learning areas as measured by the National Assessment of Educational Progress (Ahmann, 1979).

Minimum Competency Testing: Next Moves

When one considers the disagreements that exist with regard to defining and identifying competencies, developing and selecting instruments to measure minimum levels, and establishing minimum standards, it is not surprising that there is widespread debate with respect to the major directions the minimum competency testing movement should follow. At least three points of view are frequently expressed.

The first strongly supports the notion that minimum competency testing is an idea whose time has come (National Assessment of

Educational Progress, 1978). As stated by Ebel, such testing will not only benefit education, but also society. This opinion is based on the following points:

1. Excellence in learning should be a primary goal of every teacher and every school.
2. Schools and teachers should be held accountable to provide good environments in which pupils can learn.
3. Schools should be concerned primarily about cognitive learning.
4. Schools in different locations should pursue substantially similar objectives.
5. The degree of achievement of every educational goal is measurable, and such measurements are essential for effective education.

A second point of view holds that the basic motives behind the minimum competency testing movement are clearly political (Glass and Smith, 1978). A battle for control of the schools is underway. Although the public does not wish to tell the professional educator how to conduct education business, it does wish to establish its right to specify what that business will produce. Setting minimum competency standards is important in the contest to control the schools, but is meaningless when viewed as a technical, scientific endeavor.

The third point of view is an effort to escape from an impossible dilemma. On the one hand, the public believes that the present system of awarding diplomas at the end of secondary education is meaningless, and that minimum competency testing will improve matters. On the other hand, the minimum competency testing movement, which is focused on a fictional line separating the "competent" and the "incompetent," is a carnival of horrors to measurement specialists (Page, 1978). Dichotomies such as "competent" and "incompetent" are irrational. Education has long outgrown such a categorical view of human abilities. Available evidence supports the assumption that student abilities are normally distributed, for all practical purposes; this should be recognized in any effort to measure the outcomes of the educational process.

One way of doing this is to design a diploma that includes the results of a subtest-centered achievement test battery covering the most important school learning areas. Such a diploma would be called "scaled certification" (Page, 1978). The battery might have only four or five subscales (language skills, reading comprehension, mathematics, science and social studies) and would be norm-referenced. No specific interpretations of the test scores would be made by educators, leaving to others the interpretations that are appropriate. These decision-makers will learn soon enough which levels of achievement in the various learning areas hold promise in terms of performance in adult life or in further schooling. Some would be

highly relevant while others would not.

No matter which point of view one supports, it is clear that the time has come to introduce much more realism into the debate. Doing so means that this focus on reality should permeate the entire educational enterprise, beginning with the fundamental question of the nature and scope its goals. Consider the following.

Through the years educators seem to have been all too willing to believe that schools serve as the major agency to solve virtually all of our problems, whether cognitive, affective, or psychomotor in nature. In other words, we have repeatedly promised to do a great deal for a great many. These pledges were often made with eagerness and, sadly, without any particular thought about the skills and resources needed to accomplish them. Meanwhile, the public has largely accepted our promises and, unsurprisingly, expects results.

The results have not been forthcoming; huge credibility gaps now exist, and they are surfacing in the argument about minimum competencies. In keeping with past claims, the public wants the "whole" child educated for successful adulthood, evidently the entire span of adulthood. Is this possible within the elementary and secondary school years? Certainly not.

The entire idea suggests some kind of educational system with a massive "front end" load. Somehow each graduate from high school will have mastered all essential competencies at a level sufficient for him or her to interact successfully with the adult environment in terms of socially-ascribed, self-selected, and self-developed roles. At age 18, each will be at least minimally equipped to handle an adulthood spanning 50 or more years. This is most unlikely, to say the least.

A realistic view of the situation will certainly include an effort to rearrange and focus the educational effort. Some goals (usually affective) are hardly teachable within a school setting, and some learning areas (citizenship, for instance) are not highly school-specific. In these cases schools can help, but that's all. This notion needs to be emphasized. On the other hand, other goals (usually cognitive) are quite teachable, and other learning areas (mathematics, foreign languages, etc.) are highly school-specific. Here the role of the schools is dominant, and this too needs to be emphasized.

If schools stress what they do best and test accordingly, the intensity of many of the problems associated with minimum competency testing will be reduced. Such testing would reflect only a part of the family of competencies needed in adulthood, that part which is of primary concern to the elementary and secondary schools. Broadening the scope of the program to include other areas should be resisted. Also to be stoutly resisted is any identification of cutoff scores and rigid, unbending interpretations of the terms "competent" and "incompetent."

Our task is to improve the quality of tests by concentrating on a somewhat modest number of highly significant competencies that subsume lesser skills, and by describing them fully through detailed test specifications (Popham, 1978). In other words, the cornerstone of such a testing program should be objective-centered tests (both paper-and-pencil and performance tests as needed) based on learning areas that are heavily cognitive and school-specific, without cutoff levels that supposedly represent minimum competence. This approach is built on realistic goals, permits us to be more careful, and ultimately will improve the quality of minimum competency testing.

References

Ahmann, J. S. Differential changes in levels of achievement for students in three age groups. *Educational Studies*, 1979, *10*, (in press).

Ahmann, J. S. and Glock, M. D. *Evaluating pupil growth: Principles of tests and measurement* (5th ed.). Boston: Allyn and Bacon, 1975.

Block, J. H. The "C" in CBE. *Educational Researcher*, 1978 *7*, 13-16.

Bossone, R. M. *Minimum competencies: A national survey*. New York: Graduate School and University Center, City University of New York, 1978.

Brickell, H. M. Seven key notes on minimum competency testing. *Phi Delta Kappan*, 1978, *59*, 589-592.

Clearinghouse for Applied Performance Testing. Setting standards in minimum competency testing. *Newsletter, Clearinghouse for Applied Performance Testing*, 1978, *3*(5), 4-6.

Glass, G. V. Matthew Arnold and minimum competence. *Educational Forum*, 1978, *42*, 139-144.

Glass, G. V., and Smith, M. L. The technology and politics of standards. *Educational Technology*, 1978, *23*, 12-18.

Millman, J. *Strategies for constructing criterion-referenced assessment instruments*. Paper presented at the NAEP Large-Scale Assessment Conference, Boulder, CO, June 1978.

National Assessment of Educational Progress. *Guide for an assessment of basic life skills*. Denver: Author, 1977.

National Assessment of Educational Progress. *Checkup: A national assessment of health awareness among 17-year-olds and young adults*. Denver: Author, 1978a.

National Assessment of Educational Progress. "Testing" debate launches conference. *NAEP Newsletter*, 1978b, *11*(4), 3, 5.

Page, E. B. Escaping the minimum competency dilemma: Scaled certification. *Newsletter, Clearinghouse for Applied Performance Testing*, 1978, *3*(5), 6-12.

Popham, W. J. Practical criterion-referenced measures for intrastate

evaluation. *Educational Technology*, 1978, *23*, 19-23.

Slater, S. J., and Nafziger, D. H. Alternate approaches to paper-and-pencil testing using applied performance measures. *Newsletter, Clearinghouse for Applied Performance Testing*, 1978, *3*(6), 1-8.

Spady, W. G. Competency based education: A bandwagon in search of a definition. *Educational Researcher*, 1977, *6*, 9-14.

Zieky, M. J. and Livingston, S. A. *Manual for setting standards on basic skills assessment tests*. Princeton, N.J.: Educational Testing Service, 1977.

NSPER: 78

Portland, Oregon

WALTER E. HATHAWAY
Portland, Oregon Public Schools

W. R. NANCE
Oregon State Department of Education

GILBERT SAX
University of Washington

COMPETENCY MEASUREMENT AT THE LOCAL LEVEL: A CASE STUDY OF THE PORTLAND (OREGON) PUBLIC SCHOOLS

Walter E. Hathaway

Introduction

In accordance with a state mandate and in response to requests from parents, teachers and administrators, the Portland School District has developed a new Rasch based city-wide minimum competency achievement testing program. The program's grade 4 through 8 reading and mathematics tests were the first to be developed and were administered district-wide for the first time in fall 1977. Grade 3 reading and mathematics tests and grade 3 through 8 language usage tests are being field tested this fall. Reseach and resource development efforts have been mounted to prepare the way for the extension of the program to the secondary level and to other content areas such as writing, science and social studies.

The people responsible for the learning of Portland students felt a minimum competency achievement testing program, having the following characteristics, was needed:

It should help students become competent beginning as early as the third grade rather than denying diplomas to students who had arrived at the late stages of their secondary education without acquiring required competencies.

It should focus on basic skills competencies, initially, since these were the first competencies in which students were to be certified under the state mandate and since these were the highest priorities of the school system.

It should assist teachers, parents and students in instructional decision making by giving clear, accurate and unbiased information about student performance on specific major competencies and goals within the local curriculum leading to further diagnosis and prescription in areas in which a student's progress toward meeting district graduation competencies appeared unsatisfactory.

It should assist administrators, board members and citizens in administrative and policy decisions making by giving clear, accurate and unbiased information about how Portland schools are doing in comparison with others and whether performance levels within the district are heading upward, remaining constant, or decreasing over time.

It should give every student a chance to be measured at his or her individual functional level and thus avoid discouraging some students while failing to challenge others.

There should be as little testing as possible and each test required should be short enough to be administered in approximately one 50 minute period.

Building upon four years of cooperative work with the Tri-County Goal Development Project and with the Northwest Evaluation Association (NWEA) and utilizing the Rasch model, Portland teachers, curriculum specialists and evaluators have worked together to build such a program.

It has been almost 20 years since Georg Rasch first proposed his logistic model for improving school measurement (Rasch 1960, 1961, 1966a, 1966b). The Portland School System's application of this model to minimum competency achievement testing suggests that it leads to better basic skills measurement and reporting than is possible from traditional approaches to test development. Some specific advantages that the Rasch approach offers to minimum competency achievement testing include:

Competencies and items can be allowed to vary over classrooms, buildings, districts and time as instructional and curricular priorities vary and yet valid cross sectional and longitudinal comparisons can be made. Thus, for the first time, the widening gap between the instruction and management needs to which basic skills testing must respond can be closed.

Equivalent forms of a test can be readily generated.

Item "difficulty" level can be specified with respect to an underlying curriculum scale rather than some sample of students.

Equal interval measurement of basic skills performance rather than mere rank order measurement is possible.

Student progress in a subject area and on individual competencies can be directly related to the curriculum of the subject and the domain of the competency as well as to the performance of norm groups.

Error of measurement is more precisely determined and is related to each score on a test.

The advantages of both norm referenced and criterion referenced approaches to basic skills measurement can simultaneously be realized by a single testing program.

These advantages of application of the Rasch model to minimum competency achievement measurement stem from the following

properties of the Rasch model and Rasch scaling:

> The Rasch estimate of the level of an item does not vary over different samples of students (sample free item calibration) and,
>
> The Rasch estimate of a student's achievement level does not vary over different tests (item free person measurement).

The minimum competency achievement tests developed by the Portland School System using the Rasch Model have two principal distinctive features: first, they employ a Rasch scaling system that permits reporting student progress toward mastery of competencies in terms of development on an interval curriculum continuum as well as in relation to norm groups; and second, the tests are designed to be efficiently administered to students at their level of functioning.

This paper describes the competencies measured, the process used in developing the tests, the test themselves, the distinctive features of the program, and the resultant reports and their uses.

1. What Competencies?

The Oregon State Board of Education's Graduation Requirement Regulations identify ten areas in which local districts are to define competencies to be used as a basis for certifying students for high school graduation. Traditional attendance and course completion requirements are also mandated. The ten competency areas are listed in Figure 1. The Portland School Board worked together with the community to develop thirty-four competencies spanning the ten state competency areas. These thirty-four competencies are listed in Appendix 1. The classes of 1978 through 1980 must be certified in the first three state competency areas only. Competencies in the remaining seven areas in addition to the first three are to be certified beginning with the class of 1981.

Neither state regulations nor district policy require that certification of competencies be based upon tests. The current district procedures for competency certification involve the assignment of adopted competencies to required high school courses. The attainment of these competencies is determined for each student by teachers as part of the process of evaluating a student's course work and is certified by the granting of a passing grade in the course to which a competency is assigned.

The Portland minimum competency testing program has been created to identify and help students who are not making satisfactory progress toward attaining and being certified on required minimum competencies. This identification and assistance process begins with

Figure 1

Areas in Which the Oregon State Board of Education's Regulations Require Local Districts to Define Competencies

By Class of 1978	1.	Read, Write, Speak, Listen;
	2.	Analyze;
	3.	Compute;
By Class of 1981	4.	Use basic scientific and technological processes;
	5.	Develop and maintain a healthy mind and body;
	6.	Be an informed citizen in the community, state, and nation;
	7.	Be an informed citizen in interaction with the environment;
	8.	Be an informed citizen on the streets and highways;
	9.	Be an informed consumer of goods and services;
	10.	Function within an occupation or continue education leading to a career.

the third grade and stretches up through the eighth grade. Pilot testing is being conducted on the effects of using similar tests and test data as one of a number of determinants employed by high school teachers in certifying their students on competencies.

The Rasch Calibrated Item Bank Based Test Development Process

Each of the series of tests in reading, mathematics, and language usage was constructed according to blueprints in which the major objectively measurable district basic skill competencies were identified along with subcompetencies selected by teacher committees. The number of items needed to represent each competency is also found on the blueprints (see Figures 2, 3, and 4).

During the four years preceding the development of the minimum competency achievement tests, Portland and other member districts of the Northwest Evaluation Association had carried on intensive field testing of items in reading, mathematics, and language usage. Rasch analysis and linking procedures were employed to create a bank of goal and competency referenced test items that were Rasch calibrated and scaled and thus had known item levels. The intent of this activity was to make it possible to draw from the bank items of appropriate identified levels and with acceptable item statistics related to the goals and competencies to be measured, and thus to develop the capability to engage in precision development of a flexible and yet valid and reliable testing program at a minimum cost in dollars and time.

Most of the items used on the Portland tests have been drawn from the Northwest Evaluation Association's Rasch Calibrated Item Banks. These banks now contain approximately 1,400 Rasch screened, calibrated and linked items in each of the subject areas of mathematics, reading and language usage. The Portland tests were constructed by selecting items from the bank that represented the competencies in the blueprints, and by writing, field testing and Rasching new items where insufficient items were present.

Many of the competencies in the test blueprints prepared by Portland curriculum and evaluation specialists were completely covered by Rasch calibrated NWEA bank items of appropriate level and of proven quality (based on Rasch and traditional item analysis date). The initial bank calibration of items accomplished through field testing was redone on the larger groups of students taking the tests in the Portland program. This provided the district as well as the NWEA bank with firmer and more reliable calibrations for future use. Also, as Portland filled out its test blueprints by writing items for competencies not adequately covered by the bank, additional items were calibrated and made available to the bank. As other

Figure 2

Grades 4 through 8 — Mathematics Test Plan
Portland Minimum Competency Testing Program

Rasch Unit (RIT) Scale 155							250
Rasch Item Levles Spanned by Each Test	**158—177**	**171—190**	**184—203**	**197—216**	**210—229**	**223—242**	**236—255**
Competency	*# Items on Test Level 817*	*# Items on Test Level 818*	*# Items on Test Level 819*	*# Items on Test Level 820*	*# Items on Test Level 821*	*# Items on Test Level 822*	*# Items on Test Level 823*
1. The student can add whole numbers.	5	5	5	0	0	0	0
2. The student can subtract whole numbers.	5	5	5	5	0	0	0
3. The student can multiply whole numbers.	0	5	5	5	5	0	0
4. The student can divide whole numbers.	0	5	5	5	5	5	0
5. The student can order, compare, rename and represent whole numbers.	5	5	5	5	5	5	7

6. The student can order, compare, rename and represent frational numbers (fractions, decimals, percents).	0	5	5	5	5	5	5	
7. The student can compute with fractions.	0	0	0	5	5	5	5	
8. The student can compute with decimals and percents.	0	0	0	5	5	5	5	
9. The student can use knowledge of geometry.	5	5	5	5	5	5	5	
10. The student can use knowledge of measurement.	5	5	5	5	5	5	5	
11. The student can interpret and use graphs, statistics and probability.	5	5	5	5	5	5	8	
12. The student can solve story (word) problems.	5	5	5	5	5	5	5	
13. The student can use the strategies and processes of problem solving.	5	5	5	0	0	5	5	
TOTAL of Items	40	55	55	55	50	50	50	97

Figure 3

Grades 4 through 8 — Reading Test Plan
Portland Achievement Level Tests

Rasch Unit (RIT) Scale	155					255
Rasch Item Levels Spanned by Each Test		**158—177**	**171—190**	**184—203**	**197—216**	**210—229**	**223—242**
Competency		*# Items on Test Level 317*	*# Items on Test Level 318*	*# Items on Test Level 319*	*# Items on Test Level 320*	*# Items on Test Level 321*	*# Items on Test Level 322*
1. The student can interpret meanings of commonly used words. (Context Clues; Synonyms, Antonyms; Structure Components; Multiple Meanings)		14	16	12	12	11	11
2. The student can comprehend the literal meanings or explicit content of written materials. (Recall of Details, Interpreting Directions; Sequence of Details; Classification of Facts; Recall Stated Mail Idea)		16	16	17	16	14	11

3. The student can interpret implied and related meanings from the content and presentation of written mterials. (Drawing Inferences; Recognizing Cause and Effect; Prediction of Events; Summary & Synthesis)	10	7	10	9	10	12
4. The student can evalute the intent, validity and worth of written materials. (Fact and Opinion; Merit, Accuracy and Persuasion; Internal and External Validity; Conclusion & Resolution; Bias and Underlying Assumptions)	5	6	6	8	10	11
TOTAL of Items	45	45	45	45	45	45

Figure 4

Language Usage Field Test Plan
Portland Achievement Level Tests
Grades 3 through 8

Test Level	1	2	3	4	5	6	7	8
Approx. Grade Placement of Student	Low 3	Low 4 Avg. 3	Low 5 Avg. 4 High 3	Low 6 Avg. 5 High 4	Low 7 Avg. 6 High 5	Low 8 Avg. 7 High 6	Low 9 Avg. 8 High 7	Low 10 Avg. 9 High 8
Approx. Level Expressed in RIT's and G.E.'s	RIT 145—164 1.5—2.5	RIT 158—177 2.5—3.5	RIT 171—190 3.5—4.5	RIT 184—203 4.5—5.5	RIT 197—216 5.5—6.5	RIT 210—229 6.5—7.5	RIT 223—242 7.5—8.5	RIT 236—255 8.5—9.5
Number of Items for: Competency								
1. The student knows important facts and principles regarding the nature and functions of language.	5	5	5	5	5	5	5	5

2. The student can recognize and use fundamental sentence and paragraph structures.	6	6	6	6	6	6	6	6
3. The student can use basic grammar correctly	6	6	6	6	6	6	6	6
4. The student can punctuate correctly.	6	6	6	6	6	6	6	6
5. The student can capitalize correctly.	6	6	6	6	6	6	6	6
6. The student can spell correctly.	15	15	15	15	15	15	15	15
7. The student can recognize and use the elements of effective written expression (e.g., coherence, clarity, economy, and consistency).	6	6	6	6	6	6	6	6

8. The student knows the characteristics and uses of different kinds of writing (e.g., narration, summary, poetry, etc.)	5	5	5	5	5	5	5	5
TOTAL Items for Test	55	55	55	55	55	55	55	55

districts use the bank in similar fashion, they derive the same benefits and are able to make similar contributions, thus leading to a stronger and more useful resource for test development for all NWEA members and eventually for the profession as a whole.

3. Who Determines How to Measure the Competencies?

The Portland Minimum Competency Achievement Tests were constructed by the district's central evaluation department and by area evaluation administrators in collaboration with central and area curriculum personnel and teachers. Policies and procedures affecting the program are the continuing responsibility of the central-area evaluation committee and the joint curriculum-evaluation committee. These groups will continue to refine and improve the tests and will explore possibilities for measurement at other levels and in other areas of the curriculum over the next two years. All major policy decisions affecting the program were approved by the superintendent and the board of education. Selection of the competencies to be measured was carried out by committees of teachers recommended by each of the three areas under the direction of a curriculum specialist agreed on by the three areas with the resultant recommended competencies being submitted for board approval. Teachers and curriculum specialists from each of the three district administrative areas were similarly involved in selecting and writing items needed to cover the test blueprints.

The Tests

There are currently seven levels of reading tests and seven levels of mathematics tests in the Portland grade 4 through 8 minimum competency testing system. These test levels are matched to students not on the basis of their grade level, but on the basis of their approximate performance level within the subject area. The match of test to student is determined by previous test results or by short locator tests for students new to the system. This functional assignment of levels permits the measurement of each student's achievement accurately but in only a single approximately 50 minute long test period for each subject. The relative brevity of the tests makes possible both fall and spring testing in all selected subject areas. This makes available the instructional and administrative advantages of testing at each of these times.

The Portland Minimum Competency Achievement Tests were designed to measure the achievement of students in reading and mathematics accurately, efficiently and in accord with the competencies and priorities established in the district's curriculum. Such

resources as the district scope and sequence charts in mathematics and reading were used by district teachers and other curriculum specialists as they worked with evaluation personnel to identify and develop test items in the appropriate numbers and at the correct levels. Similar procedures have been used in developing the grade 3 reading and mathematics tests and the language test battery being field tested this fall.

Rasch Scale Based Reporting

The first principal distinctive feature of the Portland Minimum Competency Achievement Tests is that their results are reported in terms of an equal interval scale based upon Rasch analysis of test items representing a range of achievement in each subject extending from the simplest to the most complex levels of performance on the curriculum continuum for that subject. The Rasch scale is independent of a norming sample. It is arrived at by a procedure that partials out the effect of ability of groups performing on an item, leaving an estimate of the true relative scale level of the item compared with that of other items with which it is associated in a test. Through Rasch analysis the effect of ability on item difficulty is eliminated and through the process of linking tests together by overlapping items it is possible to equate these item difficulties to a single Rasch scale for each subject.

It should be noted that by combining the Rasch scaling feature with the competency referencing feature, the flexibility exists to substitute items or to add or eliminate competencies and items from the tests while maintaining comparability of the measures secured from year to year. This feature is not usually available with traditional norm-referenced measurement systems.

Functional Administration

The second distinctive feature of the Portland Minimum Competency Achievement Tests is that they are designed to make it possible for a student of any level of ability to take a test on which he or she can perform successfully and thus be measured efficiently. Placement of students on the tests is accomplished by an analysis of their performance at their last previous testing. The Rasch achievement estimates secured from these tests permit placement of students in the following administration of the testing program at a level at which most students are able to perform successfully on approximately 60 percent of the items. Short Rasch calibrated locator tests have been designed for placement of students new to the system. By placing students accurately in tests on which they can successfully respond to approximately 60 percent of the items a high degree of

reliability of measurement can be secured with fewer items than are contained in wide range achievement tests. This is because more items are functioning to measure what the student can do, especially for high and low achieving students. By designing the tests and related teacher manuals carefully, it is possible to test all students in a class at a single session even though they are taking different levels of tests appropriate to their level of functioning and even in the case of multiple graded or "split" classes. An added benefit is that when retesting of absent or invalidly tested students is called for, all those to be tested in a building can be tested in a single session for each subject with one proctor.

The Reports and Their Uses

The Portland Minimum Competency Achievement Tests and their reports were designed to meet two district needs, first, to provide sound survey measurement that reflects the district's scope and sequence in order to support administrative decision making; and second, to provide competency referenced measurement of student progress toward mastery of individual minimum competencies needed for high school graduation in order to support the instructional decisions needed to lead students toward competency.

A special committee of principals and teachers was formed to advise the test developers on the testing program as a whole, and on the reporting of results in particular. Students and parents were also consulted. Basically, two forms of reporting have been used to date. The first shows for every student whether or not the criterion for each minimum competency has been passed. It also identifies all students who did not reach the criterion on each competency. This information is provided both by class and by student (see Figures 5 and 6). These reports have been found to be very valuable to teachers in working with students and parents to help students toward the certification of high school graduation competencies which, by district policy, is done at the secondary level using test results as only one among several other factors. If this report indicates a student may have problems in an area of required competency, the teachers are encouraged to do further diagnostic testing and inquiry. If a problem area is confirmed and further pinpointed through these efforts, then such measures are used to alleviate the problem as: assignment to individualized instruction in building learning resource centers, assignment of school and community tutors, grouping of students with common difficulties for work with teachers and aides, and development of programs of home instruction with parents and tutors.

Research is being conducted on the identification of the most appropriate criterion levels for "passing" each competency. In the interim the district is using a criterion of the Rasch scale equivalent of

Figure 5

School 100 **Lakeside** **Teacher: John Jones** **Grade 3**

Fall 1978 Reading Goal Reort

	Goal	Passing	Students Possibly Needing Additional Work on this Goal	
1.	The student can interpret meanings of commonly used words. (Word Mng.)	10	Ralph Loren	Mathew
2.	The student can comprehend the literal meanings or explicit content of written materials. (Lit. Comp.)	11	Ralph	Roxanne
3.	The student can interpret implied and related meanings from the content and presentation of written materials. (Int. Comp.)	10	Ralph Dennis	Roxanne
4.	The student can evaluate the intent, validity and worth of written materials. (Eval. Comp.)	7	Tanya Eve Ralph	Dennis Roxanne Mathew

Figure 6

School 100 **Lakeside** **Teacher: John Jones** **Grade 3**

Fall 1978 Reading Goal Report

Student	Word Mng.	Lit. Comp.	Int. Comp.	Eval. Comp.
Caroline	Pass	Pass	Pass	Pass
Chris	Pass	Pass	Pass	Pass
Tanya	Pass	Pass	Pass	
Eve	Pass	Pass	Pass	
Ralph*				
Loren	Pass	Pass	Pass	
Chelsea	Pass	Pass	Pass	Pass
Saui	Pass	Pass	Pass	Pass
Dennis	Pass	Pass		
Roxanne*				
Mathew*	Pass	Pass	Pass	
John	Pass	Pass	Pass	Pass
Greg	Pass	Pass	Pass	Pass

*Indicates minimum standards target student.

Figure 7

Fall 1977 Achievement Level Test Student Report

Rank	Mathematics	Rit Scale:	170	178	185	193	200	207	215	222	230
					M = 40				M = 60		
Student Name		M Score	RIT	Very Low	Low	Low Average	Average	High Average	High	Very High	
Meredith		65	223							X	
Nathan		63	219						X		
Missy		62	214						X		
Heather		61	216						X		
Randy		59	214					X			
Darcy		59	214					X			
Scott		59	213					X			
Betty		56	209					X			
Parrish		56	209					X			

| | | |

Figure 8

Fall 1977 Achievement Level Test Student Report (Alphabetic)

Alpha Mathematics RIT Scale: 170 178 185 193 200 207 215 222 230

M = 40 M = 60

Student Name	M Score	RIT	Very Low	Low	Low Average	Average	High Average	High	Very High
Ronald	48	197				X			
Kelly	50	200				X			
Ronald	53	204					X		
Adrienn	45	193				X			
Donald	41	187			X				
Chrissy	33	174	X						
Lance	39	184		X					
Adam	44	191			X				
Nathan	63	219						X	
Mike	47	196				X			

Tiffany	43	190	X
Eric	39	184	X
Parrish	56	209	X
Meredith	65	223	X
Bryce	51	201	X
Mark	53	204	X
Heather	61	216	X
Missy	62	218	X
Demetriu	47	195	X
Randy	59	214	X
Scott	59	213	X
Din	43	189	X
Betty	56	209	X
Michelle	48	197	X
Darcy	59	214	X
Charles	49	198	X
Lonnie	45	192	X

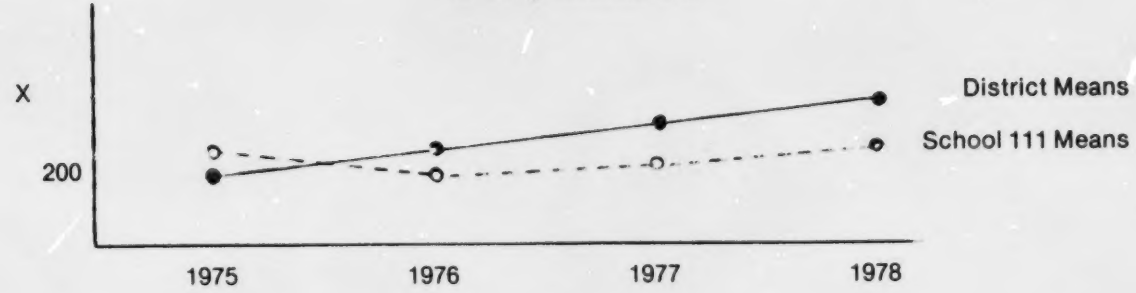
* = score based onthe competency progress test

111

Figure 9

Longitudinal School Progess Report

Grade 4
School 111 Means vs. District Means
Reading Achievement



one standard deviation below the district grade level mean for the competency at the first districtwide administration of each test.

A second type of report provided to each teacher shows the overall subject achievement position of every child in the class in relation to a distribution curve for the grade level to which the student belongs. This information is provided both alphabetically and by rank order (see Figures 7 and 8).

In addition to plotting each student's position under the normal curve, this second set of reports gives for each student the Rasch achievement score and the standard score (mean of 50, standard deviation of 10). The standard score is based on the performance of all students in the grade level group to which the student belongs. These reports have been found valuable for, among other things, grouping students and program evaluation.

In addition to the reports provided to teachers, student competency and subject achievement is reported in other forms and at different levels of aggregation to assist administrative decision making. For example, reports giving student rank within grade rather than class are provided to principals who use them for such things as Title I and gifted student identification and parent consultation. And grade, building, area and district means and standard deviations are reported longitudinally by subject to assist administrators in school program management. A sample of one of the current administrative reports is provided in Figure 9. Regression approaches to comparing actual performance to that predictable by school and ability factors are being studied.

Conclusion

While the effort described in this paper has been exciting and the results gratifying, the path to Rasch based minimum competency testing has not been an easy one to travel. Mounting such a program has been more complicated and difficult than purchasing available instruments or settling for simple, locally developed criterion measures. The planning and development challenge associated with this test development program has placed great demands on the energies and ingenuity of the Portland School District staff. It has required coordination of such disparate capabilities as computer programming, item writing, test formatting, test blueprinting, development of scoring and reporting formats, in-service education, development of support manuals for principals, teachers, and test coordinators, editing and review of items for race and sex bias, Rasch analysis, linking, field testing, test scoring, answer sheet design, development of placement procedures for functional level testing, statistical analysis of item information to screen out poor items, and test packaging and distribution. Nevertheless, the results

of the program and its benefits seem to more than justify the effort that has gone into its development.

The Portland School System looks forward to continued sharing of the results of their efforts in the area of Rasch based minimum competency achievement testing and reporting with other school systems. It is our hope that now that some initial paths have been broken to the benefits which can come from this approach, its advantages may become more readilty and generally available.

Appendix 1

Minimum Competencies
Portland Public Schools

1. The student is able to listen effectively.
2. The student is able to speak effectively.
3. The student is able to view effectively.
4. The student is able to read effectively.
5. The student is able to write effectively.
6. The student is able to apply basic computation skills to everyday problems.
7. The student is able to make reasonably accurate mental estimates of the solutions to basic arithmetic problems.
8. The student is able to apply basic measuring skills.
9. The student is able to order and compare numbers.
10. The student is able to present and explain numerical information.
11. The student is able to use common mathematical formulas.
12. The student is able to recognize and work with simple geometric figures.
13. The student is able to identify uses of calculators and computers.
14. The student understnds that matter has structure and is found in various states.
15. The student has a basic understanding of energy, its nature, its limits, and its uses.
16. The student knows that machines extend the physical capacities of human beings.
17. The student has a basic understanding of the inter-relationships between people and the earth and its natural resources.
18. The student has a basic scientific understanding of the varieties of living organisms.
19. The student has a basic understanding of the relationships among various organisms.
20. The student is able to use the basic principles of the scientific method in simple problem solving.
21. The student knows the meaning and importance of one's self concept.
22. The student has a basic knowledge of human growth and development.
23. The student understands the importance of physical and mental health.
24. The student is able to participate in leisure activities.

25. The student has the knowledge and skills necessary for safe living.
26. The student knows about basic human resources available in the community.
27. The student is able to function effectively and responsibly as a citizen.
28. The student has knowledge and skills necessary for a steward of the environment.
29. The student is able to function as a responsible individual on the streets and highways.
30. The student is able to function responsibly as a producer and consumer of goods and services within the American economic system.
31. The student is able to plan personal finances effectively.
32. The student knows of the various kinds of taxes and their uses and the obligation of citizens to support the structures of local, state and federal governments.
33. The student knows and is able to use effectively information related to career choices.
34. The student possesses and is able to use basic skills necessary to obtain and keep a job.

References

Anderson, J., Kearney, G.E., and Everett, A.V. An evaluation of Rasch's structural model for test items. *The British Journal of Mathematical and Statistical Psychology*, 1968, 21, 231-238.

Bailey, Stephen, et al. "Improving Educational Achievemnet," National Academy of Education, April 1978.

Brooks, R.D. *An empirical investigation of the Rasch ratio-scale model for item difficulty indexes.* Doctoral dissertation, University of Iowa, Ann Arbor, Michigan: University microfilms, 1965, No. 65-434.

Clark, James P. and Thompson, Scott. Competency Tests and Graduation Requirements, National Association of Secondary School Principals, Reston, Virginia. 1976

Gadway, Charles I. and Wilson, H.A. "Functional Literacy Basic Reading Highlights," National Right to Read Effort and National Assessment of Educational Progress, Denver, Colorado. 1976.

Hathaway, W.E. "The State of the Competency Testing Movement," *ETS Basic Assessment News Letter*, Spring, 1978.

Hathaway, W.E. "Competency Categories and Competencies in CBE Programs," AERA Regional Conference on Competency and Control in Education, Portland, Oregon. October 1977.

Hathaway, W.E. "Competency Based Education: The Oregon Approach," American Association for School Administrators Annual Convention, Las Vegas, Nevada. February 1977.

Hathaway, W.E. "Secondary Education/Competency Based Education: The Oregon Experience," ACT Conference on Competency Based Education, Eugene, Oregon. February 1977.

Hathaway, W.E. "Competency Based Education: Definitional Issues and Implications for Cooperation Across Educational Institutions and Level," Council of Chief State School Officers Study Commission, Competency Based Education Conference, Denver, Colorado. July 1976.

Hathaway, W.E. *Educational Reform: The Role of Competency Based Education* (editor and contributor). Proceedings of a conference sponsored by the State Commission of the Council of Chief State School Officers. Northwest Regional Educational Laboratory, Portland, Oregon. July 1976.

Hathaway, W.E. "Decentralized Minimum High School Graduation Requirements in the Context of Competency Based Education — The Oregon Approach," The 6th Annual Conference on Large-Scale Assessment, Boulder, Colorado. June 1976.

Hathaway, W.E. "A Model for the Evaluation of Statewide Educational Reform: The Oregon Competency Based Education Program," American Educational Research Association, San Francisco, California. April 1976.

Northcutt, Norval. *Adult Performance Level Study Report*, University of Texas, Austin, Texas. 1975.

Rasch, G. *Probabilistic models for some intelligence and attainment tests.*. Copenhagen: Danish Institute for Educational Research, 1960.

Rasch, G. On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley Symposium on mathemtical statistics*. Berkeley: University of California Press, 1961, IV, 321-334.

Rasch, G. An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology*, 1966, 19, 49-57. (a)

Rasch, G. An individualistic approach to item analysis. In P.F. Lazarsfeld And N.W. Henry (Eds.), *Readings in mathematics social science*. Chicago: Science Research Associates, 1966, pp. 89-108. (b)

Tinsley, H.E.A. An investigation of the Rasch simple logistic model for tests of intelligence or attainment. Unpublished docotral dissertation, University of Minnesota, 1971.

Whitely, S.E. and Dawis, R.V. The nature of objectivity with the Rasch Model. *Journal of Educational Measurement*, Vol. II, No. 2, Fall 1974, pp. 163-178.

Wright, B. Sample-free test calibration and person measurement. *Proceedings of the 1967 Invitational Conference on Testing Problems*. Princeton, N.J.: Educational Testing Service, 1967. pp. 85-101.

Wright, B. and Panchapakesan, N. A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 1969, 29, 23-48.

Wright, B.D. Solving measurement problems with the Rasch Model. *Journal of Educational Measurement*, Vol. 14, No. 2, Summer 1977. pp. 97-116.

Notes

Note: The author gratefully acknowledges the contributions of numerous colleagues to this paper and, more importantly, their leadership and participation in the work the paper describes. Special acknowledgement is made to the following members of the Portland Public School's Evaluation Department; Victor Doherty, Dean Forbes, Fred Forster, George Ingebo, Jim Holmes, Ron Houser and Bill Matson.

THE ASSESSMENT OF COMPETENCE
(An Oregon Case Study)

W. R. Nance

Oregon's graduating classes of 1978 were the first to complete the secondary school program under requirements which included verification of competency achievement in specified areas.

Events which preceded this landmark in Oregon education included: 1) an assessment of educational needs by the State Department of Education and a private marketing and research firm; 2) a series of town hall meetings involving local publics, the Superintendent of Public Instruction (SPI), and the State Board of Education (SBE); 3) two drafts of proposed graduation requirements developed by educators, students, and various publics; 4) five public hearings at meetings of the SBE; 5) adoption of the requirements in September, 1972 by the SBE; 6) guidelines developed by six pilot school districts through the efforts of some 300 educators and community members and subsequently synthesized by a team of project members and State Department of Education staff.

Implementation of the program began with the freshman class of 1974. Between that time and graduation in 1978, modification of the program has occurred. Improved, more realistic guidelines have been developed. Some time extension has been granted to local districts and graduation competencies have been redefined as "student performance representing demonstrable ability to *apply* knowledge, understanding, and/or skills to success in life role functions."

In response to the questions posed for this symposium, the answers from the Oregon experience follow:

1. What Competencies?

Students in the graduating classes of 1978-1980 must demonstrate competence in the areas of reading, writing, listening, speaking, analyzing, and computing. The class of 1981 will have the following areas added: using basic scientific and technological processes; developing and maintaining a healthy mind and body; being an informed citizen in the community, state, and nation; interacting with the environment; being an informed citizen on the streets and highways; being an informed consumer of goods and services; functioning within a career or continuing education.

2. Who Selects the Competencies to be Measured?

The SBE establishes the areas in which competency must be verified. Each local school board is required to adopt the specific competency statements, including number and level of difficulty.

3. Who Sets the Minimums?

Each local board of education.

4. Who Determines How To Measure Them?

Local district professional staff, usually with assistance from state or intermediate education service district consultants make these decisions.

5. How to Measure?

In Oregon a variety of measurement techniques are used, actual student demonstration, simulation activities and criterion-referenced tests. It must be remembered that in the Oregon program competence requires the ability to *apply* skills, knowledge, and/or understandings, not merely to possess them.

6. When to Measure?

Times are designated by the local school district during the elementary-secondary program. The only state requirement is that verification of attainment of all locally adopted competencies occur prior to a student's receiving a diploma.

7. One Minimum or many?

In Oregon most districts set a single set of minimums and then are encouraged to provide for unique needs and abilities of individual students by altering performance indicators, substituting competencies, and/or waiving competencies.

8. How High the Minimum?

Oregon school districts are encouraged to develop competencies at a level that the local community through its board of education is willing to accept as evidence that the schools have provided students with the basic minimum ability needed to function in those life roles identified by the SBE, i.e., as an individual, learner, citizen, pro-

ducer, consumer, and family member. As might be expected levels of difficulty vary from district to district, but typical levels of difficulty are: be able to read a 200 word article in a newspaper and answer 4 out of 5 questions correctly; be able to read a standard apartment rental contract and state three conditions which either party to the agreement must fulfill; be able to read a job application form and provide the required information in writing; be able to balance and maintain a checkbook over a specified period of time and a specified number of deposits and withdrawals; be able to determine the added costs associated with time payment purchases; and be able to compose a workable budget that includes a minimum of six items.

9. Minimums for Students or Schools?

Minimums are established for students. Lest the minimum competencies become the focus of the educational program and thereby become the maximums also, Oregon requires that a planned course statement be written for all courses in grades 9-12. The statements must include a content overview, clearly stated course goals in terms of learner outcomes, and any competencies assigned to that course for verification.

10. What to do With the Incompetent?

First, identify those students who may have or are having difficulty achieving the required competencies. Their problem may be a physical, sensory, or other handicap that will necessitate adjustment of the competency requirements or a specialized instructional program. Second, give the students every bit of assistance available. Oregon schools are committed to this and are providing a variety of additional services to students, such as competency centers, tutoring, special summer sessions, and staff on duty before and after regular school hours.

For those students who cannot meet the attendance, unit of credit, and/or competency requirements, the state board of education recommends issuance of a certificate of competence by the local district that identifies those requirements which a student has completed.

COMPETENCY BASED EDUCATION

Gilbert Sax

Some 17 or 18 years ago, as an assistant professor of educational psychology at the University of Hawaii, I was asked by the Hawaii Department of Education to explain to teachers how they could best implement a new report card system. Instead of the more traditional reporting systems, the department of education was recommending the use of a plus, check and minus to indicate, respectively, working up to ability, below ability or far below ability. My suggestion, which was never taken seriously, was for every teacher to give every student a minus and to continue to do so in perpetuity. My thinking, if simple, was not simple-minded. After all, I explained, no one works up to his or her ability at any time or for any reason (as you read these words, for example, are you eating a completely balanced meal, exercising and writing your responses to this paper for publication? If not, you aren't working up to your own ability either). The only exceptions I could think of might include neonates and the dead, both groups, by their very nature, being excluded from the topic of competency based education.

The next time I became involved with competency based education (CBE) was some 7 or 8 years ago at the University of Washington. It was a summer session, and I was teaching an introductory statistics class. I was concerned about the math anxiety expressed by some of the students, and I was determined to do something about it. I decided to give every student a pretest so that no one would be surprised or shocked at the nature of the items that would appear on the actual tests. For example, if there were 3 items on the test that required students to compute a mean from different kinds of distributions, three items of the same type would appear on the pretest. Then, I set 90-100 percent as an A grade; 80-90 percent was a B. Any score lower than 80 percent would require that the student retake the examination after receiving some remediation. Any student who retook an examination could not earn an A on that test; still, there was no way to receive a grade lower than B in the course.

Does this sound familiar? With the exception that the stakes may be higher for the public school student than they were for the students in my class and that my students could repeat the tests as often as they liked, I had developed a rather typical CBE program.

I'll never do it again, and I should have known better even at that time. Let me explain why. First, anxiety was not reduced one whit. Indeed, I soon discovered that students were quite willing to pay for an education they might not receive by cheating on the exams. The

reason was the arbitrary 80 or 90 percent. A score of 79 percent was as bad as 10 percent and an 80 percent was as good as 100 percent for those who retook quizzes. Under the traditional grading systems, a graduate student would receive full credit for a C as long as that grade could be balanced by an A in some other course.

A second reason for dropping my "mastery" program is that it didn't necessarily measure mastery. For example, I administered a final examination to this class and discovered that they had little difficulty in forgetting what they had just mastered. What is mastered today may be flustered tomorrow.

A third reason for my disillusionment with mastery concerns a real student whose alias is Albert. Albert did reasonably well on the first pretest, but failed the quiz itself. He also failed the second, third, fourth, fifth, sixth, and seventh make-up exams for the first quiz. Now don't tell me Albert was unmotivated! Believe me, he worked to pass that exam, and I would have done anything to help him pass including whispering the correct answers to him if I were only sure that he wouldn't copy them incorrectly. We both eventually called it quits, and he received a passing grade during another quarter with another member of the faculty. The message is clear: if you don't have the time, materials, or funds to provide for remediation, forget CBE programs. They won't work.

A fourth reason for my concern with "proficiency" programs is more psychometrically oriented. Albert complained, for example, that the items on the pretest and on the regular test and make-up exams were not equivalent in difficulty. And, at least on the pretest and regular examination, he was right! I intended to make them equal in difficulty but, as is true on most teacher-made tests, I did not have the advantage of pretesting all items. I am quite sure that the seventh make-up test bore only a superficial resemblance to the pretest or any of the other tests administered to Albert. Unless the tests are comparable, 80 percent on one test will not necessarily equal 80 percent on another.

I must admit to another problem about which the passing students cared nothing and the failing students never let me forget, the percentages required for mastery were completely arbitrary, subjective, and without one shred of evidence to support their use. This can be tolerated (but not defended) in those situations where the consequences are minor; it cannot be tolerated when a diploma will be withheld. Standards are not established by frozen subjectivity but by accumulating data that demonstrate their validity for a specified purpose, for a specified group of students, and for a given group of items having known parameters.

Any model of instruction requires some consideration of objectives, some idea of what is to be accomplished. Teachers should not be asked to hold themselves accountable for reaching any objective

unless they are also given the requisite time, materials, and methods. Long-term or ultimate objectives are not under the school's control, and teachers should not be held responsible for their attainment.

What minimum competencies should every student have? I am not convinced that the ones I have seen are particularly "basic." Here is my list to consider:

1. Learning how to prevent pregnancy
2. Learning to select a reasonable diet
3. Learning who or what agencies can answer any other questions that may be asked.

I doubt that most schools will want to discuss pregnancy and contraception with fifth, sixth, and seventh grade students when it will do the most good. I further doubt that knowing what a good diet is will in any way affect student eating habits. My third objective, however, is much more basic. If Uncle Carl is identified as one who can answer a given question, I'd give the student credit, provided, of course, that the answer is correct.

Henry Brickell, in the May 1978 edition of the *Phi Delta Kappan*, provides an example of what he means by a "life skills" test item: To saw very hard metal, should you buy a hacksaw blade with few teeth or many teeth? I hope that among the correct answers could be found, "I wouldn't cut very hard metal," "I would borrow two blades and try them," and "I would ask my son-in-law who teaches industrial education classes."

The answer to the question "what competencies are basic?" is crucial. I am not convinced that it is necessary to be able to "use highway and city maps" (assuming you can still find one), an objective taken from Florida's Functional Literary Skills Test. I can read maps and I still got lost in every major European city while I was on sabbatical leave. The problem is that you can't use a map unless you know where you are. Other objectives such as "Determine the main idea inferred from a selection" and "Determine the cause and effect of an action" seem to depend upon which "selection" or "action" one has in mind. Some selections are quite simple and the determination of cause and effect relations can be extremely difficult.

I have some difficulty believing that students *must* know how to estimate "the amount of simple interest." The Truth in Lending Law makes that objective passe. Besides, if enough of us become confused on a given topic, we can demand that our state legislatures require unit pricing in the supermarket and other similar laws to make life more bearable and education less demanding.

Another reason why I don't believe that we will agree to a set of basic competencies is that we have never put our philosophical houses in order. To the philosophical pragmatists, "content" and "method" are partly used to solve problems of concern to students. To realists, such as Harry Broudy, the school is not life but a

preparation for it. All content is basic to pragmatists if that content can be used to solve problems; the school's responsibility is to teach students how to solve problems. Any content that cannot serve this function would not be taught. Yet, despite the accusation that American schools are "Dewey-eyed" I have seen no objectives that demand that students learn how to solve problems *important to them*. Harry Broudy once told me that he jokingly suggested on one occasion that the schools replace teachers by psychologists, psychiatrists, and social workers and that the information-giving function of the school be replaced by "adjustatoriums." Much to Harry's dismay, he received numerous requests from administrators asking how this change could be implemented!

The schools are not adjustatoriums; nor are they simple information centers.

The American school system reminds me of a cafeteria where students can select from a wide variety of offerings. It may not serve everyone equally well, but it ought not try to guarantee that everyone become a vegetarian. In its attempt to specify what all students must eat, when they will eat and what the state will do to them if they prefer a different fare, competency based education will, I predict, be dumped along with the other garbage we have been forced to swallow or have been too timid or complacent to refuse.

Certainly we want all students to be well-nourished "mentally" as well as "physically." All students for example, should learn decoding and other reading skills to whatever levels they are capable of reaching within the time, monetary and interest limitations of the schools, the children, and their parents. But it is just as asinine to demand that every student learn some arbitrary list of required content as it is to demand that nourishment can only be attained if every student eats carrots.

A second component of instruction requires that teachers consider the entry behaviors of their students. It has become almost trite to state that children differ in innumerable ways: intelligence, socioeconomic levels, attitudes, personality, and motivation, and no one seriously denies this variability, no one except those who are hell-bent in what they see as educational reform and what I see as a revival of the worst aspects of American education.

I have taught retarded children and children who were considered to be "slow learners." Many of these children liked school and worked as hard as I could reasonably expect. Yet, some of these pupils could not possibly pass the tests developed by those who advocate CBE programs. Their reward for their efforts will be a certificate of attendance while those who learn easily and need school the least will receive the more cherished diploma. Once again we've created an intellectual elite. Mr. Gary Hart's* statement in the May

*Mr. Hart introduced the California Proficiency Law to that state's Assembly.

issues of the *Phi Delta Kappan* that "There is a danger that the weight of any sanction will fall unfairly upon the student" gives me little solace in this regard.

I have no qualms, however, about attempts to diagnose learning difficulties and attempts to provide remedial services. Neither of these activities requires any "minimum level of proficiency," whatever one means by that phrase. Nor would I be unwilling to emphasize reading, writing, and arithmetic over other subjects, provided that arbitrary and capricious standards and consequences not be enforced by legislative or school board fiat.

Admittedly, many junior and senior high school students are unwilling to learn despite the effort of parents, teachers, counselors, psychologists, and a host of other professional workers. Some are in trouble constantly and make it difficult for other, more serious students to learn. The possibility of not receiving a diploma may encourage a few of these students to take schooling more seriously. But the vast majority of these students are likely to become even more disruptive than they are now under pressure.

Again, the solution is not CBE. In the state of Washington, many schools do not even take roll. Students may come in late to class (if they come at all), they disrupt teaching, and there are no consequences. Educators are rightfully reluctant to impose physical punishment, and to send kids home serves only to reinforce the negative behavior. What, then, can the schools do except to threaten to withhold diplomas?

Let me suggest another possibility. After all attempts to gain the student's cooperation have failed, the student should be expelled permanently from school no matter what that child's chronological age. Students who do not want to learn should not be permitted to disrupt education for everyone else. It is true that these students will not find jobs, will create even a greater social liability outside of school than in, and that labor and parents will oppose this option. These consequences are all true, but irrelevant. The students did not have jobs in school, and considering the disruption caused by constant attempts at "acting out," the benefit to the school, other children, and for the student himself or herself will be well worth it. Certainly such a drastic action as expulsion should require every legal protection of the student. My point is only that CBE is likely to create more problems than it will solve and that there are alternatives that legislatures and school boards should consider.

A third element that enters into any instructional model involves the selection of an instructional strategy. Although many CBE programs require remediation, remediation does not depend upon CBE programs. Educators who want the benefits of remediation need not accept CBE also.

Somehow the advocates of criterion-referenced tests have persuaded legislators and school district administrators that good in-

struction depends upon those kinds of tests. Let me explain my general opposition to criterion-referenced tests:

1. They encourage preposterous statements. For example, Gary Hart says that the intent of the California Pupil Proficiency Law is to "measure the progress of each student in mastering basic skills rather than the pupil's performance relative to his or her classmates." In the very next sentence Mr. Hart contradicts what he just said, "To establish an arbitrary passing score based on a norm can fail to insure mastery by all pupils (when the majority of pupils are deficient in the basics) or doom some students to failure (when all students have mastered the skills, with insignificant differences in pupil performance)."

What Mr. Hart does not understand is that nothing is as arbitrary as the cutoff score used on criterion-referenced tests. A mark of 70 percent (or any other value) derives its meaning from a standardization group and does not have the same meaning on different tests. A 70 percent on an easy test is not equivalent to a 70 percent on a difficult examination. If new items are constructed each year to measure "mastery," passing or failing will depend upon item selection decisions and not upon student knowledge.

2. They are psychometrically crude. Norm-referenced tests are often accused by their critics of being an artificial means of spreading people out. In contrast, the criterion-referenced test supposedly measures the amount of knowledge students have.

The truth of the matter, however, is that criterion-referenced tests are required to spread students out at a single point, the cutoff score. If the cutoff point is 70 percent (whatever that means), the ideal criterion-referenced test will place passing students above that point, while placing failing students well below it. Any knowledgeable student (in a norm-referenced sense) who, because of errors of measurement, falls below the cutoff score, has been misplaced. Equally misplaced is the student who "lucks out" and passes.

3. In setting standards, a normative interpretation is usually implied. Is 70 percent good or not? It is good if the test items are so difficult that only a few persons do that well; it is poor if no one does that poorly. Without a normative set of standards, any may is meaningless.

4. Methods of test construction and evaluation are well known for norm-referenced tests, whereas the technology for criterion-referenced tests is currently in the Paleozoic stage. When students fail examinations and have diplomas withheld, one consequence is lawsuits. Any notion of fair play will require an examination of the examination. Be prepared to answer such questions as (a) on what bases were items selected? (b) what principles were used to select one item and not another? (c) what evidence is there that the cutoff score

is reasonable and fair? (d) what are the psychometric properties of each item? (e) what are the reliability and validity coefficients? and (f) what is the error of measurement?

A fourth and final element of an instructional model is evaluation or decision-making. The question to be answered is "What decisions are to be made following the testing of students?" Again, I would argue against a policy that would deny a high school diploma to students who are judged by the school and by the parents to be doing reasonable work. If students are still deficient in basic skills at the end of the twelfth grade, they can be encouraged to continue studying in remedial classes or at community colleges. If they are still deficient at the end of the thirteenth year, it is time again to reassess the student's goals and the ability of the school to provide for the kind of achievement the student and his or her parents demand.

I can understand the public's demand for accountability in education, but I am always amazed when teachers, administrators, and school boards accede to the demands. After all, I know of no profession that holds its clients accountable for their failure to live up to arbitrary standards established by the profession. Indeed, no profession even holds itself accountable for the *consequences* of its practice. All that is required is that professional *practice* be reasonable. Lawyers do not have a minimum number of cases they must win to continue practicing; similarly, the medical profession holds no one responsible should a patient die unless there has been malpractice on the part of the physician.

A profession does, however, have an obligation to the public to select its practitioners with care. Rather than testing the teachers and administrators, we have found it more palatable to test the students. Instead, I would suggest that we administer tests to all teachers to make sure that they know their subject matter, are able to read or write, and can perform "basic" operations in arithmetic. But fair is fair, let's give them the same opportunities to pass (or fail) as we give the students!

NSPER: 78

Discussion and Recommendations

Discussion and Recommendations

A major element of each NSPER session is the opportunity for participants to work together in small groups to expand their understanding of the topic being considered. Based on this understanding each small group is charged with the responsibility of making recommendations appropriate to the nature and complexity of the issues that are highlighted in the formal paper presentations or in group discussions. The topic of minimum competency testing, as discussed within the framework of the ten organizing questions, provoked intense conversation that resulted in a number of concerns being raised repeatedly across the work groups. Some of these will be presented here. This discussion will be followed by a summary of the responses to the ten questions across the three sessions and finally by a list of the recommendations made.

General Discussion

Some time was spent in each work group attempting to gain perspective on the origins of the apparently sudden rush to minimum competency testing programs which has occurred in the last two or three years. Several precipitating factors were noted. Among these: a generalized vague public dissatisfaction with the schools; the current press towards accountability in all sectors of public life; and the traditional reluctance of teachers and schools to inform the public about the job that they have been doing. Having little direct testimony from the schools and evidence of poor preparation of some students, the public has concluded that the schools are failing. In effect the schools are being blamed for student failures.

Group discussion also centered on a few basic issues preparatory to dealing with the questions posed at the outset of each session. Some attention was devoted to what is meant by competence. Is it achievement? One answer given was competency is a basic, demonstrable, measurable skill, when competency is defined in an educational context. However a great deal of uncertainty remained about the definition of competency in terms of level of skill and range of skills. If competencies are defined in terms of occupations or life skills, how viable are they? How long should a student be certified as competent?

Concern was expressed that minimum competency testing could lead to teaching to the test; to narrowing or re-focusing of the curriculum to insure that students could be certified as minimally competent but concurrently lowering the achievement ceiling for all students. Given the speed with which new ideas have been adopted

and then abandoned in public education in the past, the opinion that minimum competency testing could also be a passing fad was expressed by several.

On the positive side, minimum competency testing was seen as providing a structure for formal identification of students who were having difficulties in some subject areas. Such identification could lead to special attention to the learning difficulties experienced by these students through re-teaching. It was also seen as a vehicle for emphasis on staff development in areas such as writing objectives, learning the various uses of test information, or interpretation of research for use in the classroom.

In sum, however, the participants seemed to feel that minimum competency testing had been accepted too rapidly before the proper definition of goals and objectives for such an effort could have been completed. Actual testing had commenced with less than adequate preparation in some locales and the effects of such programs on students were only beginning to be addressed.

Responses to the Ten Structuring Questions

In addition to the general discussion reported above, each work group focused attention on the ten structuring questions. Greater emphasis was placed on some questions rather than others in all groups. Some tended to focus their attention on a subset of the questions. The responses of each work group will not be detailed here. Summaries of the responses to each question will be made.

1. What Competencies? The answer that came out most consistently was school skills, that is, the three R's. It was pointed out that overlap does exist between school and life competencies. There was also agreement that any competency selected should have the characteristics of observability and stability across instances. This suggested that school skills should receive the most attention, especially since there were concerns about the accuracy with which competencies could be related to various occupations. Questions were also raised about the appropriateness of using a single set of competencies for all populations that might be tested.

The answer to the question, what competencies, came out fairly unambiguiously: school skills, reading, writing, and arithmetic.

2. Who selects the competencies to be measured? There were two schools of thought in regard to this question. One of these advocated maximum feasible participation. It was suggested that the various publics be surveyed about their specific concerns in terms of what should be tested. People at all levels should be involved thus providing a wide base of support and involvement in selecting competencies.

The other view that received some support was that the state

board should set the competencies. A slightly different position was that evaluators and practitioners, selected as advisors to state agency officials, should make these decisions.

In essence two answers to this question were generated by the work groups. One advocated full community participation in selection of competencies to be tested while the other favored reliance on established authority and expertise in the field of education.

3. Who sets the minimums? This question apparently received less attention than most. Those who did consider it advocated that a committee of evaluators and practitioners should set the minimums. However, it was also stipulated that there should be continuous reassessment of the appropriateness of the minimums set.

4. Who determines how to measure? There was some consensus across the three sessions that measurement specialists ought to have the primary responsibility in deciding the best approach to testing. It was felt, however, that teachers should have some advisory role.

5. How to measure? The range of responses to this question was quite broad. At one end of the spectrum was an unwillingness to take any position until the objectives or items to be tested were specified. Others felt that existing norm referenced tests should be bought and used. A third response was that both norm referenced and specially constructed competency tests should be used. Finally, one group felt that a national test, with state and local additions where necessary, would be the best approach.

6. When to measure? Here again there seems to have been some consensus. The importance of testing early in the educational career was stressed fairly uniformly along with a recommendation that testing be conducted frequently. It was felt that the results of these tests should be used for diagnosis of student difficulties and reteaching. it was also suggested by at least one group that the results of minimum competency tests could be used to assess school programs with future improvement in mind.

7. One minimum or many? There seemed to be some agreement that all minimums should be established in terms of the individual student.

8. How high the minimum? It was stated in response to this question that the minimum level is arbitrary. And, in recognition of that fact, group processes should be employed to determine the level of performance desired.

9. Minimums for students or schools? The answer to this question came out fairly clearly under question seven. Minimums should be established for students, not for schools.

10. What to do with the incompetent? Participants felt that students who fail the tests should be provided with remediation. They should also have multiple opportunities to retake the test. At least one group felt that promotion should not be tied to passing a

minimum competency test. None felt that it should be the sole basis for withholding the diploma.

Recommendations

1. Where minimum competency testing has been mandated, responsible educators should proceed cautiously, allowing up to three years for pilot activities. Existing minimum competency testing programs should be assessed across several dimensions, including, process, use of results, difficulties encountered, educational implications, and unintended effects. The results of such assessments should serve as a guiding base for the development of new programs.

2. The ten questions that served to structure the NSPER: 78 sessions should be examined and answered by any locale that is considering implementing a minimum competency program before it is initiated.

3. When planning the development and implementation of competency assessment programs, care should be taken that attention is focused on the instructional program, insuring that it includes appropriate content. In developing any competency test, care should be taken to insure that the test reflects the instructional program and does not diverge to the extent that substantial curriculum revisions are required.

4. Schools should retain existing assessment programs and monitor achievement through both norm referenced and competency tests. Exchanging one form of assessment for another would result in disruption of existing data bases, making it difficult to maintain adequate information about student performance.

5. Alternatives to minimum competency testing programs should be identified, described and evaluated.

6. A critical review of existing research and other literature on minimum competency testing programs should be completed. This would provide a base that could be used by various education agencies that are faced with the development and implementation of a minimum competency testing program. It would also be a synthesis of existing knowledge that could provide insights into many aspects of the issue.

7. The bases of the public push for minimum competency testing should be more carefully assessed. The public should be integrally involved in the implementation of any minimum competency testing program.